



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 179453

TO: Konstantina Katcheves

Location: 2a60 / 2c70

Art Unit: 1636

Friday, February 17, 2006

Case Serial Number: 10/627007

From: Noble Jarrell

Location: Biotech-Chem Library

Rem 1B71

Phone: 272-2556

Noble.jarrell@uspto.gov

### Search Notes

*This Page Blank (uspto)*

179453

**STIC-Biotech/ChemLib**

---

**From:** Chan, Christina  
**Sent:** Monday, February 13, 2006 3:47 PM  
**To:** Katcheves, Konstantina; STIC-Biotech/ChemLib  
**Subject:** RE: 10-627007

Please rush. Thanks Chris

Chris Chan  
TC 1600 New Hire Training Coordinator and SPE 1644  
(571)-272-0841  
Remsen, 3E89

-----Original Message-----

**From:** Katcheves, Konstantina  
**Sent:** Monday, February 13, 2006 3:05 PM  
**To:** Chan, Christina  
**Subject:** 10-627007

Christina:

Would you approve a search of SEQ ID NO:5

Thanks,  
Tina

**Konstantina Katcheves**  
**Patent Examiner, AU1636**  
**Phone: (571) 272-0768**  
**Room: REM 2A60**  
**Mail: REM 2C70**

2/13/2006

Noted  
For 2/17/05  
INA  
5PR Compagn  
38ML

This Page Blank (uspto)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2006, 05:20:28 ; Search time 2087 Seconds  
(without alignments)  
2062.488 Million cell updates/sec

Title: US-10-627-007-5

Perfect score: 92

Sequence: 1 cggatcaggagaccctcg.....cagccgggtgctggcttttt 92

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_est3.\*  
4: gb\_est4.\*  
5: gb\_est5.\*  
6: gb\_est6.\*  
7: gb\_est7.\*  
8: gb\_est8.\*  
9: gb\_est9.\*  
10: gb\_est10.\*  
11: gb\_est11.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	100.0	625	10	CW179236
2	92	100.0	878	10	CL672134
3	76.8	83.5	775	10	CL664024
4	64	69.6	242	2	BG895158
5	47.2	51.3	1082	9	AF094862
6	39	42.4	904	10	CNS01GKM
7	30.2	32.8	853	2	BF676283
8	29.8	32.4	755	10	AG414181
9	29	31.5	976	11	CNS05RUK
10	28.8	31.3	469	9	AZ087480
11	28.8	31.3	615	6	CD728108
12	28.8	31.3	744	5	BU231483
13	28.8	31.3	746	5	BU365131
14	28.8	31.3	774	5	BU427039
15	28.8	31.3	780	6	CF256576
16	28.8	31.3	791	6	CF254599
17	28.8	31.3	804	5	BU478133
18	28.8	31.3	1272	9	BZ579155
19	28.6	31.1	345	1	AV209814
20	28.6	31.1	572	8	CX956602
21	28.6	31.1	572	8	CX956603
22	28.6	31.1	579	3	BI344684

23	28.6	31.1	681	7	CR568140
24	28.6	31.1	698	7	CR568141
25	28.6	31.1	727	8	DR878616
26	28.6	31.1	727	8	DR878617
27	28.6	31.1	779	3	BI333728
28	28.4	30.9	741	10	CG900138
29	28.2	30.7	372	4	AK203556
30	28.2	30.7	739	2	BE544496
31	28	30.4	669	1	AI322522
32	27.8	30.2	594	9	AZ624975
33	27.8	30.2	754	9	AZ894224
34	27.8	30.2	768	11	CR163471
35	27.8	30.2	888	6	CF592897
36	27.6	30.0	240	1	AA125488
37	27.6	30.0	411	1	AI596810
38	27.6	30.0	601	6	CA883090
39	27.6	30.0	638	6	CD804829
40	27.6	30.0	672	1	BB185328
41	27.6	30.0	692	5	BY728178
42	27.6	30.0	1037	8	DR128951
43	27.6	30.0	1038	4	AK021089
44	27.6	30.0	1586	4	AK039322
45	27.4	29.8	388	1	AL908917

#### ALIGNMENTS

RESULT 1  
CW179236  
LOCUS

DEFINITION

104 593 11159994 116 36626 075 Sorghum bicolor genomic clone 11159994, genomic survey  
(LibID: 104) Sorghum bicolor genomic clone 11159994, genomic survey  
sequence.

ACCESSION  
CW179236

VERSION  
GSS.

KEYWORDS  
GSS.

SOURCE  
Sorghum bicolor (sorghum)

ORGANISM  
Sorghum bicolor

REFERENCE  
Bedell J.A., Budiman M.A., Nunberg A., Citek R.W., Robbins D.,

AUTHORS  
Jones J., Flick E., Rohlfing T., Fries J., Bradford K., Rabinowicz P.D., Lakey N., McCombie W.R., Jeddeloh J.A. and Martienssen R.A.

TITLE  
Sorghum genome sequencing by methylation filtration

JOURNAL  
PLOS Biol. 3 (1), e13 (2005)

COMMENT  
15660154

CONTACT: Bedell JA

ORION GENOMICS, LLC

4041 Forest Park Ave, St. Louis, MO 63108, USA

Tel: 314 615 6979

Fax: 314 615 5975

Email: jbedell@oriongenomics.com

Seq primer: T3 Reverse

Class: methylation filtered

High quality sequence stop: 625.

Location/Qualifiers

1..625

/organism="Sorghum bicolor"

/mol\_type="genomic DNA"

/cultivar="ATx623"

/db\_xref="taxon:4558"

/clone="11159994"

/clone\_lib="Sorghum methylation filtered library (LibID: 104)"

/notes="Organ: leaf; Vector: pBCSK(-); Site: 1: HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5

kb fraction, ligated into HincII-digested pBSK(-) vector and electroporated into E. coli cells. This is a methylation filtered library."

ORIGIN

Query Match 100.0%; Score 92; DB 10; Length 625;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-21;  
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCGATCAGGAGACCTCGCGAGAACCTGAAAGCAGCACATTGCTCATTGCTTCCA 60  
 |||  
 Db 12 CGCGATCAGGAGACCTCGCGAGAACCTGAAAGCAGCACATTGCTCATTGCTTCCA 71  
 |||

QY 61 GTATTACTTAGCAGCGGGTGTGCTTTT 92  
 |||  
 Db 72 GTATTACTTAGCAGCGGGTGTGCTTTT 103  
 |||

## RESULT 2

CL672134/c  
 LOCUS  
 DEFINITION  
 PRI016a.D04 - PRI016a.B21 (878) Mixed stage fosmid library of P.  
 pacificus var. California Pristionchus pacificus genomic, genomic  
 survey sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

CL672134.1 GI:50172033  
 GSS.  
 Pristionchus pacificus  
 Pristionchus pacificus  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
 Neodiplogasteridae; Pristionchus.

REFERENCE  
 AUTHORS  
 TITLE

JOURNAL  
 PUBMED

COMMENT

Contact: Sommer RJ  
 Evolutionary Biology  
 Max-Planck-Institute for Developmental Biology  
 Spemannstr. 37-39, Tuebingen D-72076, Germany  
 Tel: 00497071601371  
 Fax: 00497071601498  
 Email: ralf.sommer@tuebingen.mpg.de  
 This library was generated at Caltech, Pasadena, USA and end  
 sequenced at Vancouver, Canada.  
 Seq primer: T7  
 Class: fosmid ends.

## FEATURES

source

1..878  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus  
 var. California"  
 /note="Vector: pEpifos-5 Fosmid vector"

## ORIGIN

Query Match 100.0%; Score 92; DB 10; Length 878;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-21;  
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCGATCAGGAGACCTCGCGAGAACCTGAAAGCAGCACATTGCTCATTGCTTCCA 60  
 |||  
 Db 294 CGCGATCAGGAGACCTCGCGAGAACCTGAAAGCAGCACATTGCTCATTGCTTCCA 235  
 |||

QY 61 GTATTACTTAGCAGCGGGTGTGCTTTT 92  
 |||  
 Db 234 GTATTACTTAGCAGCGGGTGTGCTTTT 203  
 |||

## RESULT 3

CL664024/c  
 LOCUS  
 DEFINITION

PRI0146a.A06 - PRI0146a.B21 (775) Mixed stage fosmid library of P.  
 pacificus var. California Pristionchus pacificus genomic, genomic  
 survey sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

CL664024.1 GI:50153989  
 GSS.  
 Pristionchus pacificus  
 Pristionchus pacificus  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;  
 Neodiplogasteridae; Pristionchus.

REFERENCE  
 AUTHORS  
 TITLE

JOURNAL  
 PUBMED

COMMENT

Contact: Sommer RJ  
 Evolutionary Biology  
 Max-Planck-Institute for Developmental Biology  
 Spemannstr. 37-39, Tuebingen D-72076, Germany  
 Tel: 00497071601371  
 Fax: 00497071601498  
 Email: ralf.sommer@tuebingen.mpg.de  
 This library was generated at Caltech, Pasadena, USA and end  
 sequenced at Vancouver, Canada.  
 Seq primer: T7  
 Class: fosmid ends.

## FEATURES

source

1..775  
 /organism="Pristionchus pacificus"  
 /mol\_type="genomic DNA"  
 /strain="California"  
 /db\_xref="taxon:54126"  
 /clone\_lib="Mixed stage fosmid library of P. pacificus  
 var. California"  
 /note="Vector: pEpifos-5 Fosmid vector"

## ORIGIN

Query Match 83.5%; Score 76.8; DB 10; Length 775;  
 Best Local Similarity 97.5%; Pred. No. 9.8e-16;  
 Matches 78; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCGATCAGGAGACCTCGCGAGAACCTGAAAGCAGCACATTGCTCATTGCTTCCA 60  
 |||  
 Db 296 CGCGATCAGGAGACCTCGCGAGAACCTGAAAGCAGCACATTGCTCATTGCTTCCA 237  
 |||

QY 61 GTATTACTTAGCAGCGGG 80  
 |||  
 Db 236 GTATTACTTAGCAGTCCGG 217  
 |||

## RESULT 4

BG895158/c

LOCUS

DEFINITION  
 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.

ACCESSION  
 VERSION  
 EST.

KEYWORDS

SOURCE

ORGANISM

BG895158 242 bp mRNA linear EST 05-JUN-2001  
 358371 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.  
 BG895158  
 BG895158.1 GI:14305399  
 EST.  
 Sus scrofa (pig)  
 Sus scrofa  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;  
 Sus.

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

1 (bases 1 to 242)  
 Fahrenkrug,S.C., Smith,T.P.L., Fraking,B.A., Cho,J., White,J.,  
 Valliet,J., Wise,T., Rohrer,G.A., Pertea,G., Sultana,R.,  
 Quackenbush,J. and Keele,J.W.  
 Porcine gene discovery by normalized cDNA-library sequencing and  
 EST cluster assembly  
 Mamm. Genome 13 (8), 475-478 (2002)  
 1226715

## COMMENT

Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@mail.marc.usda.gov  
 Single pass sequencing. Bases called and alt trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 18  
 and -minmatch 12 options.

PCR Primers  
 FORWARD: AGGAACAGCTATGACCAT  
 BACKWARD: GTTTCCTCCAGTCAGCAGC  
 Plate: 121 row: L column: 2  
 Seq primer: ATTAGGTGACACTATAG.

## FEATURES

source

1..242  
 Location/Qualifiers  
 /organism="Sus scrofa"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9823"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /clone\_lib="MARC 1PIG"  
 /notes="Vector: pCMV SPORT6; Site\_1: NotI; Site\_2: SalI;  
 Library made from pooled tissue from day 11, 13, 15, 20,  
 and 30 embryos."

## ORIGIN

Query Match 69.6%; Score 64; DB 2; Length 242;  
 Best Local Similarity 98.7%; Pred. No. 2.3e-11;  
 Matches 75; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGCGATCAGAACACCTCGCGAGAACCTGAAGCAGCAGCATTTGCTCACATTGCTT-CC 59  
 |||||  
 DB 76 CGCGATCAGAACACCTCGCGAGAACCTGAAGCAGCAGCATTTGCTCACATTGCTTNC 17  
 |||||  
 QY 60 AGTATTACTTAGCCAG 75  
 |||||  
 DV 16 AGTATTACTTAGCCAG 1

## RESULT 5

AF094862/c  
 LOCUS AF094862 Salmonella typhimurium LT2, Lambda DASH II Salmonella  
 DEFINITION typhimurium genomic clone 261-17, genomic survey sequence.  
 ACCESSION AF094862  
 VERSION AF094862.1 GI:4322704  
 KEYWORDS GSS.  
 SOURCE Salmonella typhimurium  
 ORGANISM Salmonella typhimurium  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Salmonella.

REFERENCE 1 (bases 1 to 1082)  
 AUTHORS Wong, R.M.Y. and McClelland, M.  
 TITLE End Sequences of Salmonella typhimurium LT2 Lambda DASHII Clones,  
 Li-Cor

Unpublished (1999)

CONTACT: McClelland M  
 Molecular Biology  
 Sidney Kimmel Cancer Center  
 3099 Science Park Road, San Diego, CA 92121, USA  
 Email: mclelland@lifesci.sdsu.edu  
 Class: shotgun.

## FEATURES

source

1..1082  
 Location/Qualifiers  
 /organism="Salmonella typhimurium"  
 /mol\_type="genomic DNA"  
 /strain="LT2"  
 /db\_xref="taxon:602"  
 /clone\_lib="261-T7"  
 /notes="Vector: Lambda DASH II, sequenced using Li-Cor  
 sequencer"

## ORIGIN

Query Match 51.3%; Score 47.2; DB 9; Length 1082;  
 Best Local Similarity 80.9%; Pred. No. 2.4e-05;  
 Matches 55; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
 QY 25 GAACCTGAAGCAGCAGCATTTGCTCACATTGCTTCAGTATTATTAGCCAGCCGGTGCT 84  
 |||||  
 DB 1078 GAGACTGAAGCAGCAGCATGCTCATGCTCCAGTATTATTATTTGCCAGCTTTTGCT 1019  
 |||||  
 QY 85 GGCTTTT 92  
 |||||  
 DB 1018 GGCTTTT 1011

## RESULT 6

CNS01GKM

LOCUS

DEFINITION Anopheles gambiae GSS SP6 end of clone 06P08 of NotreDamel library  
 from strain PBST of Anopheles gambiae (African malaria mosquito),  
 genomic survey sequence.

ACCESSION AL143255

VERSION AL143255.1 GI:7001417

KEYWORDS GSS.

SOURCE Anopheles gambiae (African malaria mosquito)

ORGANISM

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 Culicidae; Anophelinae; Anopheles.

REFERENCE 1 (bases 1 to 904)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :

Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)

REFERENCE 2 (bases 1 to 904)

AUTHORS Roth, C.W., Brey, P.T., Ke, Z., Collins, F.H. and Weissensbach, J.

TITLE Direct Submission

JOURNAL Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.

Roux, Paris 75015, France  
 This clone is from an A. gambiae BAC library provided by F.H.  
 Collins and sequenced by Genoscope in collaboration with the  
 Laboratory of Biochem. and Biol. Molec. of Insects, Institut  
 Pasteur.

FEATURES

source

1..904  
 Location/Qualifiers  
 /organism="Anopheles gambiae"  
 /mol\_type="genomic DNA"  
 /strain="PEST"  
 /db\_xref="taxon:7165"  
 /clone\_lib="06P08"  
 /clone\_lib="NotreDamel"  
 /note="end : SP6"

## ORIGIN

Query Match 42.4%; Score 39; DB 10; Length 904;  
 Best Local Similarity 100.0%; Pred. No. 0.017;  
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 GAAAGCAGCAGCATTTGCTCACATTGCTTCAGTATTACTT 69

DB 1 GAAAGCAGCAGCATTTGCTCACATTGCTTCAGTATTACTT 39

## RESULT 7

BF676283/c

LOCUS

DEFINITION BF676283 Homo sapiens cDNA clone IMAGE:4250842 5',  
 mRNA sequence.

ACCESSION BF676283

VERSION BF676283.1 GI:11950178

KEYWORDS EST.

SOURCE Homo sapiens (human)

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 853)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-rc@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1073 row: d column: 11
High quality sequence stop: 332.
FEATURES
Location/Qualifiers
1..853
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4250842"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 83"
/note="Organ: prostate; Vector: pDNR-LTB (Clontech);
Site 1: SfII (ggcgccgcgcgc); Site 2: SfII
(ggcctataggcc); 5' and 3' adaptor were used in cloning
as follows: 5' adaptor sequence: 5'-CAGGCCATTATGGCC-3'
and 3' adaptor sequence:
5'-ATTCTAGAGCGCGGCGCGGCACATG-DT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.4
kb (range 0.5-4.0 kb). 14/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA)."
```

```

Query Match 32.8%; Score 30.2; DB 2; Length 853;
Best Local Similarity 65.7%; Pred. No. 20;
Matches 44; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 13 GACCTCGCGGAGAACCTGAAGCAGCAGCATTTGCTCACATTGCTTCCAGTATTACTTAGC 72
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
237 GCGCTTTGTGCGAGCGGAGAGGAGGACAGTGGCCACAGCCCTTAGGGGATTCCAGGGC 178
QY 73 CAGCTGG 79
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
177 CAGCTGG 171

RESULT 8
AG414181
LOCUS Mus musculus molossinus DNA, clone:MSMg01-275N09.TU, genomic survey
DEFINITION sequence.
ACCESSION AG414181.1 GI:48057007
VERSION AG414181
KEYWORDS Mus musculus molossinus (Japanese wild mouse)
SOURCE Mus musculus molossinus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe.K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to
genomic constitution of strain C57BL/6J, as defined by BAC-end
sequence-SNP analysis
```

```

JOURNAL Homo sapiens
PUBMED Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
AUTHORS Homnidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-rc@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1073 row: d column: 11
High quality sequence stop: 332.
FEATURES
Location/Qualifiers
1..755
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-275N09.TU"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
Query Match 32.4%; Score 29.8; DB 10; Length 755;
Best Local Similarity 60.5%; Pred. No. 27;
Matches 49; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 9 GGAAGACCCCTCGCGAGAACCTGAAGCAGCAGCATTTGCTCACATTGCTTCCAGTATTACT 68
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
485 GGAAGGCTCTCTGAGAAGCAGGTGTGAGCAGCAGCAGCATCTCTTCTTCTCCAGTACATCT 544
QY 69 TAGCCAGCGCGGTCTGGCTT 89
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
545 TACCCAGCTGGTTCAGGCTT 565

RESULT 9
CNS05RUK/
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 039N23 of library A from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION ALJ51029
VERSION ALJ51029.1 GI:8244799
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.
REFERENCE 1
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizesame,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Estimation of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
PUBMED 10835645
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
```



Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.  
 Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*  
 Genome Res. 10 (7), 939-949 (2000)  
 10899143  
 3 (bases 1 to 976)  
 Genoscope.  
 Direct Submission  
 Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the *Tetraodon nigroviridis* genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES  
 source  
 1. .976  
 /organism="Tetraodon nigroviridis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:99883"  
 /clone="039N23"  
 /clone\_lib="A"  
 /notes="Genoscope sequence ID : COAA039CG12C1 end : T7"

ORIGIN  
 Query Match 31.5%; Score 29; DB 11; Length 976;  
 Best Local Similarity 63.8%; Pred. No. 55;  
 Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;  
 QY 22 GGAGAACTCAAGACGACGATGCTCACATGCTTCACATATTACTTAGCCAGCGGGT 81  
 Db 684 GCAAGGCTGACTACATCACACGCTCACAGCTGACAGTTTGTAGCTCTCATGCTGTT 625  
 QY 82 GCTGGCTTT 90  
 Db 624 GTTGCTTT 616

RESULT 10  
 AZ087480/c  
 LOCUS  
 DEFINITION  
 RPCI-23-425M13-TJB RPCI-23 Mus musculus genomic clone  
 RPCI-23-425M13, genomic survey sequence.  
 ACCESSION  
 AZ087480  
 VERSION  
 AZ087480.1 GI:7729528  
 KEYWORDS  
 GSS.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.  
 1 (bases 1 to 469)  
 REFERENCE  
 AUTHORS  
 Zhao, S., Niernman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret, B., Levins, M., Megann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.  
 Mouse BAC End Sequences from Library RPCI-23  
 Unpublished (1999)  
 Other GSSs: RPCI-23-425M13.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)

Plate: 425 row: M column: 13  
 Seq primer: SP6  
 Class: BAC ends

FEATURES  
 source  
 1. .469  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-425M13"  
 /sex="female"  
 /lab host="DH10B"  
 /clone\_lib="RPCI-23"  
 /notes="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN  
 Query Match 31.3%; Score 28.8; DB 9; Length 469;  
 Best Local Similarity 62.5%; Pred. No. 54;  
 Matches 45; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
 QY 2 GCATCAGGAGACCTCGCGAGACCTGAAGACGACGATTCGTCACATTCGTTCCAG 61  
 Db 120 GTGGTTAAGAGGACGACGCTGCTTGTAAAGACGACGATTCGTCACATTCGTTCCAG 61  
 QY 62 TATTACTTAGCC 73  
 Db 60 AATTCGCGAGCC 49

RESULT 11  
 CD728108  
 LOCUS  
 DEFINITION  
 4033546 1GAL - Chicken Intestinal Lymphocyte Gallus gallus CDNA clone 1GAL\_81021 5', mRNA sequence.  
 CD728108  
 VERSION  
 CD728108.1 GI:32278957  
 KEYWORDS  
 EST.  
 SOURCE  
 Gallus gallus (chicken)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
 1 (bases 1 to 615)  
 REFERENCE  
 AUTHORS  
 Min, W., Lillehoj, H.S., Ashwell, C.M., Matukumalli, L.K., van Tassel, C. and Han, J.Y.  
 Chicken intestinal lymphocyte EST database as a resource for the analysis of mucosal immune function  
 Unpublished (2003)  
 JOURNAL  
 COMMENT  
 Contact: Hyun S. Lillehoj  
 Animal Parasite Diseases Laboratory  
 Animal and Natural Resources Institute, USDA  
 Bldg. 1043, BARC-East, Beltsville, MD 20705, USA  
 Tel: 3015048771  
 Fax: 3015045103  
 Email: hlilleho@anri.barc.usda.gov  
 Single pass sequencing. Bases called and trimmed with phred 0.00925 using options -trim\_alt - -trim\_fasta. Vector identified by cross match using options -minmatch 12 -minscore 18  
 Plate: 81 row: O column: 21  
 Seq primer: ATTAGGTGACACTAG  
 High quality sequence stop: 615.

FEATURES  
 source  
 1. .615  
 Location/Qualifiers  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="white leghorn SC"  
 /db\_xref="taxon:9031"

/clone="IGAL\_81021"  
 /sex="mixed"  
 /tissue\_type="Gut"  
 /cell\_type="Lymphocyte"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH108"  
 /clone\_lib="IGAL - Chicken Intestinal Lymphocyte"  
 /note="Organ: Intestine; Vector: pCMV-SpORt6; Site\_1:  
 Sali; Site\_2: NotI; Normalized library from chicken gut  
 infected with coccidia duodenum and middle gut."

## ORIGIN

Query Match 31.3%; Score 28.8; DB 6; Length 615;  
 Best Local Similarity 58.0%; Pred. No. 58;  
 Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;  
 QY 4 GATCAGGAGACCCCTCGGAGAACCTGAAAGCAGCAGCATTTGCTTCACATTGCTTCCAGTA 63  
 DB 30 GGTTCGATGACCTCTGCTGAGAACATGCTACCTCTGTTTGTGTCATTTGCTGTCATGC 89  
 QY 64 TTACTTAGCCAGCCGGTCTGCTTTT 91  
 DB 90 TTATCATTGCAACTGGTCTGCTTTT 117

## RESULT 12

BU231483  
 LOCUS 744 bp mRNA linear EST 26-NOV-2002  
 DEFINITION 603948804F1 CSEQCHN23 Gallus gallus cDNA clone CHEST905h1 5', mRNA  
 sequence.  
 ACCESSION BU231483  
 VERSION BU231483.1 GI:25473276  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 1 (bases 1 to 744)  
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
 A Comprehensive Collection of Chicken cDNAs  
 Curr. Biol. 12 (22), 1965-1969 (2002)  
 12445392  
 PUMED  
 COMMENT Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES  
source

1..744  
 Location/Qualifiers  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="White Leghorn, Hixsex"  
 /db\_xref="taxon:9031"  
 /clone="CHEST905h1"  
 /dev\_stage="22"  
 /lab\_host="DH108"  
 /clone\_lib="CSEQCHN23"  
 /note="Organ: heads; Vector: pBluescript II KS(+); Site\_1:  
 EcoRI; Site\_2: NotI; This normalized library was  
 constructed from 1 million independent clones. cDNA  
 synthesis was initiated using an oligo(dT) primer, using  
 methylated C in the first strand synthesis reaction.  
 Following this first strand reaction, double-stranded cDNA  
 was bluntended, ligated to NotI adaptors, digested with  
 EcoRI, size-selected, and cloned into the NotI and EcoRI  
 compatible sites of a custom modified MCS of the  
 pBluescript (KS+) vector. The library was normalized in 2  
 rounds using conditions adapted from Soares et al., PNAS

(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
 (1996): 791, except that a significantly longer  
 reannealing hybridization was used."

## ORIGIN

Query Match 31.3%; Score 28.8; DB 5; Length 744;  
 Best Local Similarity 58.0%; Pred. No. 61;  
 Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;  
 QY 4 GATCAGGAGACCCCTCGGAGAACCTGAAAGCAGCAGCATTTGCTTCACATTGCTTCCAGTA 63  
 DB 107 GGTTCGATGACCTCTGCTGAGAACATGCTACCTCTGTTTGTGTCATTTGCTGTCATGC 166  
 QY 64 TTACTTAGCCAGCCGGTCTGCTTTT 91  
 DB 167 TTATCATTGCAACTGGTCTGCTTTT 194

## RESULT 13

BU365131  
 LOCUS 746 bp mRNA linear EST 28-NOV-2002  
 DEFINITION 603586589F1 CSEQCHN72 Gallus gallus cDNA clone CHEST543b24 5', mRNA  
 sequence.  
 ACCESSION BU365131  
 VERSION BU365131.1 GI:25873132  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 1 (bases 1 to 746)  
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
 A Comprehensive Collection of Chicken cDNAs  
 Curr. Biol. 12 (22), 1965-1969 (2002)  
 12445392  
 PUMED  
 COMMENT Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES  
source

1..746  
 Location/Qualifiers  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="Compton line 151"  
 /db\_xref="taxon:9031"  
 /clone="CHEST543b24"  
 /sex="Female"  
 /tissue\_type="cerebrum"  
 /dev\_stage="adult"  
 /lab\_host="DH108"  
 /clone\_lib="CSEQCHN72"  
 /note="Organ: brain; Vector: pBluescript II KS(+); Site\_1:  
 EcoRI; Site\_2: NotI; This normalized library was  
 constructed from 1 million independent clones. cDNA  
 synthesis was initiated using an oligo(dT) primer, using  
 methylated C in the first strand synthesis reaction.  
 Following this first strand reaction, double-stranded cDNA  
 was bluntended, ligated to NotI adaptors, digested with  
 EcoRI, size-selected, and cloned into the NotI and EcoRI  
 compatible sites of a custom modified MCS of the  
 pBluescript (KS+) vector. The library was normalized in 2  
 rounds using conditions adapted from Soares et al., PNAS  
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
 (1996): 791, except that a significantly longer  
 reannealing hybridization was used."

## ORIGIN

Query Match 31.3%; Score 28.8; DB 5; Length 746;  
 Best Local Similarity 58.0%; Pred. No. 61;  
 Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 4 GATCAGGAAGACCTCGCGGAGAACCTGAAAGCAGACATTGCTCATTGCTTCCAGTA 63  
 DB 435 GGTTTGAATGACCTGCTGGAGAACATGCTACCTCTGTTTGTGACATTGCTGCATGC 494

QY 64 TTACTTAGCCAGCGGCTGCTGCTTTT 91  
 DB 495 TTATCATTTGCACTGGTGGCTTTT 522

RESULT 14  
 BU427039 774 bp mRNA linear EST 29-NOV-2002  
 LOCUS 603954645F1 CSEQRBN09 Gallus gallus cDNA clone CHEST920112 5', mRNA  
 DEFINITION sequence.  
 ACCESSION BU427039  
 VERSION BU427039.1 GI:25919715  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 REFERENCE 1 (bases 1 to 774)  
 AUTHORS Boardman,P.E., Sans-Esquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,  
 Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.  
 TITLE A Comprehensive Collection of Chicken cDNAs  
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
 PUBMED 12445392  
 COMMENT Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES  
 source  
 1. 774  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="Layer and broiler"  
 /db\_xref="taxon:9031"  
 /clones="CHEST920112"  
 /sex="Male and female"  
 /tissue\_type="Chondrocytes isolated from growth plate cartilage"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="CSEQRBN09"  
 /note="Vector: pBluescript II KS(+); Site 1: EcoRI;  
 Site 2: NotI. This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN  
 Query Match 31.3%; Score 28.8; DB 5; Length 774;  
 Best Local Similarity 58.0%; Pred. No. 61;  
 Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 4 GATCAGGAAGACCTCGCGGAGAACCTGAAAGCAGACATTGCTCATTGCTTCCAGTA 63  
 DB 124 GGTTTGAATGACCTGCTGGAGAACATGCTACCTCTGTTTGTGACATTGCTGCATGC 183

QY 64 TTACTTAGCCAGCGGCTGCTGCTTTT 91  
 DB 184 TTATCATTTGCACTGGTGGCTTTT 211

RESULT 15  
 CF256576 780 bp mRNA linear EST 07-AUG-2003  
 LOCUS mdvnl41\_d06 Marek's disease virus-infected spleen Gallus gallus  
 DEFINITION cDNA, mRNA sequence.  
 ACCESSION CF256576  
 VERSION CF256576.1 GI:33489831  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 REFERENCE 1 (bases 1 to 780)  
 AUTHORS Wittzelli,H., Bed'Hom,B., Morin,V., Young,J.R., Whittaker,C.J.,  
 Chausse,A.M. and Zoorob,R.  
 TITLE A collection of chicken ESTs from activated immune cells  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Zoorob R  
 UPB 1983  
 CNRS  
 7 rue Guy Moquet, BP 8, 94801 Villejuif cedex, France  
 Tel: 33 1 49 58 35 00  
 Fax: 33 1 49 58 33 81  
 Email: zoorob@vjf.cnrs.fr.  
 Location/Qualifiers  
 1. 780  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9031"  
 /clone\_lib="Marek's disease virus-infected spleen"  
 /notes="Organ: Spleen; Vector: pTriplex2"

FEATURES  
 source  
 1. 780  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9031"  
 /clone\_lib="Marek's disease virus-infected spleen"  
 /notes="Organ: Spleen; Vector: pTriplex2"

ORIGIN  
 Query Match 31.3%; Score 28.8; DB 6; Length 780;  
 Best Local Similarity 58.0%; Pred. No. 61;  
 Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 4 GATCAGGAAGACCTCGCGGAGAACCTGAAAGCAGACATTGCTCATTGCTTCCAGTA 63  
 DB 631 GGTTTGAATGACCTGCTGGAGAACATGCTACCTCTGTTTGTGACATTGCTGCATGC 690

QY 64 TTACTTAGCCAGCGGCTGCTGCTTTT 91  
 DB 691 TTATCATTTGCACTGGTGGCTTTT 718

Search completed: February 16, 2006, 06:25:17  
 Job time : 2091 secs

Blank page (not for  
insertion)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2006, 04:56:28 ; Search time 338 Seconds  
(without alignments)  
1814.058 Million cell updates/sec

Title: US-10-627-007-5

Perfect score: 92

Sequence: 1 cgcgcaggaagaccctcg.....cagccgggtgctggcttttt 92

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_21:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*  
14: Geneseqn2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	92	6	ABs63103
2	45.2	49.1	60203	10	ACF65383_4
3	45.2	49.1	110000	10	ACF65385_4
4	45.2	49.1	110000	10	ACF67367_04
5	45.2	49.1	110000	10	ACF67367_33
6	45.2	49.1	110000	10	ACF67367_34
7	31	33.7	354	11	ACH94896
8	30	32.6	9603	2	AAx60894
9	28.2	30.7	1338	12	AD000018
10	28.2	30.7	1338	12	ADN98449
11	28.2	30.7	2523	5	AAx72985
12	28.2	30.7	11872	4	AAK84027
13	28.2	30.7	11872	4	AAK66893
14	27.6	30.0	1404	8	ACM43893
15	27.4	29.8	8668	13	ADR84515
16	27	29.3	2337	6	ABN95768
17	27	29.3	4669	4	ABL02136
18	26.8	29.1	558	4	AAI19231
19	26.8	29.1	558	4	ABA64237

ALIGNMENTS

RESULT 1

ABS63103  
ID ABS63103 standard; DNA; 92 BP.

XX ABS63103;

AC AC

XX 05-NOV-2002 (first entry)

DT Identification of small RNA molecules related polynucleotide #5.

DE Small RNA; sRNA; cell regulation; intercell regulation; diagnostic;

XX Small RNA; sRNA; cell regulation; intercell regulation; diagnostic;

KW antibiotic; ds.

XX Escherichia coli.

OS Escherichia coli.

XX WO200260914-A2.

 PN | 08-AUG-2002. || XX | 31-JAN-2002; 2002WO-US003147. |
XX	01-FEB-2001; 2001US-0266402P.
PR	(USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX	Gottesman S, Storz G, Repoila F, Wassarman K, Rosenow C;
XX	WPI; 2002-619223/66.
XX	New polynucleotide and polypeptide from small RNA candidates and open reading frames, useful as mediators of cell or intercell regulation and in the development of diagnostics and antibiotics.
PT	Claim 6; Page 59; 90pp; English.
PT	The invention describes an isolated Escherichia coli polynucleotide from small RNA (sRNA) candidates and open reading frames (ORFs), its complement, or its homologue. The polynucleotides or the polypeptides are useful as mediators of cell or intercell regulation. They may also be used in the development of diagnostics and antibiotics. This sequence represents a candidate polynucleotide studied for the presence of a small RNA open reading frame

XX	Sequence	92 BP; 21 A; 26 C; 23 G; 22 T; 0 U; 0 Other;
	Query Match	100.0%; Score 92; DB 6; Length 92;
	Best Local Similarity	100.0%; Pred. No. 3.3e-25;
	Matches	92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1	CGCGATCAGGAAGACCCCTCGCGGAGAACCTGAAAGCAGCAGCATTTGCTCACATTGCTTCCA 60
Db	1	CGCGATCAGGAAGACCCCTCGCGGAGAACCTGAAAGCAGCAGCATTTGCTCACATTGCTTCCA 60
Qy	61	GTATTACTTAGCCAGCCGGGTGCTGGCTTTT 92
Db	61	GTATTACTTAGCCAGCCGGGTGCTGGCTTTT 92
RESULT 2		
ACF65383_4		
	Continuation (5 of 5) of ACF65383 from base 400001 (Photorhabdus luminescens nucleotide	
	WP Sequence split into 5 fragments	LOCUS ACF65383 Accession ACF65383
	WP Fragment Name	Begin End
	WP ACF65383_0	1 110000
	WP ACF65383_1	100001 210000
	WP ACF65383_2	200001 310000
	WP ACF65383_3	300001 410000
	WP ACF65383_4	400001 460203
	Query Match	49.1%; Score 45.2; DB 10; Length 60203;
	Best Local Similarity	93.5%; Pred. No. 4e-06;
	Matches	58; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
Qy	32	AAAGCAGCAGCATTTGCTCACAATTCCTCCAGTATTACTTTAGGCAGC-CGGGTGCTGGCTTT 90
Db	14503	AAAGCAGCAGCATTTGCTCACAATTCCTCCAGTATTTTTCAGTCTGGTTCAGTCTGGTCTGGCTTT 14562
Qy	91	TT 92
Db	14563	TT 14564
RESULT 3		
ACF65385_4		
	Continuation (5 of 7) of ACF65385 from base 400001 (Photorhabdus luminescens nucleotide	
	WP Sequence split into 7 fragments	LOCUS ACF65385 Accession ACF65385
	WP Fragment Name	Begin End
	WP ACF65385_0	1 110000
	WP ACF65385_1	100001 210000
	WP ACF65385_2	200001 310000
	WP ACF65385_3	300001 410000
	WP ACF65385_4	400001 510000
	WP ACF65385_5	500001 610000
	WP ACF65385_6	600001 618776
	Query Match	49.1%; Score 45.2; DB 10; Length 110000;
	Best Local Similarity	81.2%; Pred. No. 5e-06;
	Matches	65; Conservative 0; Mismatches 13; Indels 2; Gaps 1;
Qy	15	CCCTCGCGGAGAACCTGAAAGCAGCAGCATTTGCTCAGTATTACTTCCAGTATTAC--TTTACG 72
Db	99120	CCTTGGGGGTACAGATGAAGCAGCATTTGCTCAGTATTGCTCAGTATTACTTTAGC 99179
Qy	73	CAGCGCGGTGCTGGCTTTT 92
Db	99180	CAGCCNAGTGTGGCTTCTT 99199
RESULT 4		
ACF67367_04		
	Continuation (5 of 57) of ACF67367 from base 400001 (Photorhabdus luminescens nucleotide	
	WP Sequence split into 57 fragments	LOCUS ACF67367 Accession ACF67367
	WP Fragment Name	Begin End
	WP ACF67367_00	1 110000
	WP ACF67367_01	100001 210000

WP	Fragment Name	Begin	End
WP	ACF67367_00	1	110000
WP	ACF67367_01	100001	210000
WP	ACF67367_02	200001	310000
WP	ACF67367_03	300001	410000
WP	ACF67367_04	400001	510000
WP	ACF67367_05	500001	610000
WP	ACF67367_06	600001	710000
WP	ACF67367_07	700001	810000
WP	ACF67367_08	800001	910000
WP	ACF67367_09	900001	1010000
WP	ACF67367_10	1000001	1110000
WP	ACF67367_11	1100001	1210000
WP	ACF67367_12	1200001	1310000
WP	ACF67367_13	1300001	1410000
WP	ACF67367_14	1400001	1510000
WP	ACF67367_15	1500001	1610000
WP	ACF67367_16	1600001	1710000
WP	ACF67367_17	1700001	1810000
WP	ACF67367_18	1800001	1910000
WP	ACF67367_19	1900001	2010000
WP	ACF67367_20	2000001	2110000
WP	ACF67367_21	2100001	2210000
WP	ACF67367_22	2200001	2310000
WP	ACF67367_23	2300001	2410000
WP	ACF67367_24	2400001	2510000
WP	ACF67367_25	2500001	2610000
WP	ACF67367_26	2600001	2710000
WP	ACF67367_27	2700001	2810000
WP	ACF67367_28	2800001	2910000
WP	ACF67367_29	2900001	3010000
WP	ACF67367_30	3000001	3110000
WP	ACF67367_31	3100001	3210000
WP	ACF67367_32	3200001	3310000
WP	ACF67367_33	3300001	3410000
WP	ACF67367_34	3400001	3510000
WP	ACF67367_35	3500001	3610000
WP	ACF67367_36	3600001	3710000
WP	ACF67367_37	3700001	3810000
WP	ACF67367_38	3800001	3910000
WP	ACF67367_39	3900001	4010000
WP	ACF67367_40	4000001	4110000
WP	ACF67367_41	4100001	4210000
WP	ACF67367_42	4200001	4310000
WP	ACF67367_43	4300001	4410000
WP	ACF67367_44	4400001	4510000
WP	ACF67367_45	4500001	4610000
WP	ACF67367_46	4600001	4710000
WP	ACF67367_47	4700001	4810000
WP	ACF67367_48	4800001	4910000
WP	ACF67367_49	4900001	5010000
WP	ACF67367_50	5000001	5110000
WP	ACF67367_51	5100001	5210000
WP	ACF67367_52	5200001	5310000
WP	ACF67367_53	5300001	5410000
WP	ACF67367_54	5400001	5510000
WP	ACF67367_55	5500001	5610000
WP	ACF67367_56	5600001	5648894

Query Match 49.1%; Score 45.2; DB 10; Length 110000;  
 Best Local Similarity 81.2%; Pred. No. 5e-06;  
 Matches 65; Conservative 0; Mismatches 13; Indels 2; Gaps 1;

QY	15	CCTCGCGGAGACCTGAAAGCAGCATTGCTCACATTGCTTCCAGTATTAC--TTAGC 72
DB	104254	CCCTCGGGGTACAGATGAAAGCAGCATTGCTCACATTGCTTCCAGTATTACTTTAGC 104313
QY	73	CAGCCGGGTGCTGGCTTTT 92
DB	104314	CAGCCAAGTGTGGCTTCTT 104333

RESULT 6

WP	Fragment Name	Begin	End
WP	ACF67367_00	1	110000
WP	ACF67367_01	100001	210000
WP	ACF67367_02	200001	310000
WP	ACF67367_03	300001	410000
WP	ACF67367_04	400001	510000
WP	ACF67367_05	500001	610000
WP	ACF67367_06	600001	710000
WP	ACF67367_07	700001	810000
WP	ACF67367_08	800001	910000
WP	ACF67367_09	900001	1010000
WP	ACF67367_10	1000001	1110000
WP	ACF67367_11	1100001	1210000
WP	ACF67367_12	1200001	1310000
WP	ACF67367_13	1300001	1410000
WP	ACF67367_14	1400001	1510000
WP	ACF67367_15	1500001	1610000
WP	ACF67367_16	1600001	1710000
WP	ACF67367_17	1700001	1810000
WP	ACF67367_18	1800001	1910000
WP	ACF67367_19	1900001	2010000
WP	ACF67367_20	2000001	2110000
WP	ACF67367_21	2100001	2210000
WP	ACF67367_22	2200001	2310000
WP	ACF67367_23	2300001	2410000
WP	ACF67367_24	2400001	2510000
WP	ACF67367_25	2500001	2610000
WP	ACF67367_26	2600001	2710000
WP	ACF67367_27	2700001	2810000
WP	ACF67367_28	2800001	2910000
WP	ACF67367_29	2900001	3010000
WP	ACF67367_30	3000001	3110000
WP	ACF67367_31	3100001	3210000
WP	ACF67367_32	3200001	3310000
WP	ACF67367_33	3300001	3410000
WP	ACF67367_34	3400001	3510000
WP	ACF67367_35	3500001	3610000
WP	ACF67367_36	3600001	3710000
WP	ACF67367_37	3700001	3810000
WP	ACF67367_38	3800001	3910000
WP	ACF67367_39	3900001	4010000
WP	ACF67367_40	4000001	4110000
WP	ACF67367_41	4100001	4210000
WP	ACF67367_42	4200001	4310000
WP	ACF67367_43	4300001	4410000
WP	ACF67367_44	4400001	4510000
WP	ACF67367_45	4500001	4610000
WP	ACF67367_46	4600001	4710000
WP	ACF67367_47	4700001	4810000
WP	ACF67367_48	4800001	4910000
WP	ACF67367_49	4900001	5010000
WP	ACF67367_50	5000001	5110000
WP	ACF67367_51	5100001	5210000
WP	ACF67367_52	5200001	5310000
WP	ACF67367_53	5300001	5410000
WP	ACF67367_54	5400001	5510000
WP	ACF67367_55	5500001	5610000
WP	ACF67367_56	5600001	5648894

Query Match 49.1%; Score 45.2; DB 10; Length 110000;  
 Best Local Similarity 81.2%; Pred. No. 5e-06;  
 Matches 65; Conservative 0; Mismatches 13; Indels 2; Gaps 1;

QY	15	CCTCGCGGAGACCTGAAAGCAGCATTGCTCACATTGCTTCCAGTATTAC--TTAGC 72
DB	4254	CCCTCGGGGTACAGATGAAAGCAGCATTGCTCACATTGCTTCCAGTATTACTTTAGC 4313
QY	73	CAGCCGGGTGCTGGCTTTT 92
DB	4314	CAGCCAAGTGTGGCTTCTT 4333





ischemic heart disease; ulcerative colitis.

Homo sapiens.

WO2004038003-A2.

06-MAY-2004.

24-OCT-2003; 2003WO-US033947.

25-OCT-2002; 2002US-0421061P.

25-OCT-2002; 2002US-0421080P.

25-OCT-2002; 2002US-0421552P.

30-OCT-2002; 2002US-0421614P.

30-OCT-2002; 2002US-0422177P.

30-OCT-2002; 2002US-0422178P.

15-NOV-2002; 2002US-0426355P.

15-NOV-2002; 2002US-0426384P.

15-NOV-2002; 2002US-0426394P.

15-NOV-2002; 2002US-0426430P.

15-NOV-2002; 2002US-0426916P.

27-NOV-2002; 2002US-0429224P.

27-NOV-2002; 2002US-0429275P.

27-NOV-2002; 2002US-0429302P.

27-NOV-2002; 2002US-0429326P.

27-NOV-2002; 2002US-0429651P.

04-DEC-2002; 2002US-0430645P.

04-DEC-2002; 2002US-0430651P.

04-DEC-2002; 2002US-0430657P.

04-DEC-2002; 2002US-0430663P.

04-DEC-2002; 2002US-0430668P.

05-DEC-2002; 2002US-0430937P.

05-DEC-2002; 2002US-0430965P.

05-DEC-2002; 2002US-0431458P.

12-DEC-2002; 2002US-0433251P.

12-DEC-2002; 2002US-0433500P.

13-DEC-2002; 2002US-0433316P.

13-DEC-2002; 2002US-0433318P.

23-DEC-2002; 2002US-0436236P.

03-JAN-2003; 2003US-0437914P.

17-JAN-2003; 2003US-0440820P.

17-JAN-2003; 2003US-0440821P.

18-APR-2003; 2003US-0463700P.

18-APR-2003; 2003US-0463708P.

18-APR-2003; 2003US-0463716P.

KW

XX

OS

XX

PN

XX

PD

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PR 15-NOV-2002; 2002US-0426430P.  
PR 15-NOV-2002; 2002US-0426916P.  
PR 27-NOV-2002; 2002US-0429224P.  
PR 27-NOV-2002; 2002US-0429275P.  
PR 27-NOV-2002; 2002US-0429302P.  
PR 27-NOV-2002; 2002US-0429326P.  
PR 27-NOV-2002; 2002US-0429651P.  
PR 04-DEC-2002; 2002US-0430645P.  
PR 04-DEC-2002; 2002US-0430651P.  
PR 04-DEC-2002; 2002US-0430657P.  
PR 04-DEC-2002; 2002US-0430663P.  
PR 04-DEC-2002; 2002US-0430668P.  
PR 04-DEC-2002; 2002US-0430684P.  
PR 05-DEC-2002; 2002US-0430937P.  
PR 05-DEC-2002; 2002US-0430965P.  
PR 05-DEC-2002; 2002US-0431458P.  
PR 12-DEC-2002; 2002US-0433251P.  
PR 12-DEC-2002; 2002US-0433500P.  
PR 13-DEC-2002; 2002US-0433316P.  
PR 13-DEC-2002; 2002US-0433318P.  
PR 23-DEC-2002; 2002US-0436236P.  
PR 03-JAN-2003; 2003US-0437914P.  
PR 17-JAN-2003; 2003US-0440820P.  
PR 17-JAN-2003; 2003US-0440821P.  
PR 18-APR-2003; 2003US-0463708P.  
PR 18-APR-2003; 2003US-0463716P.  
PR 18-APR-2003; 2003US-0463732P.  
PR 02-MAY-2003; 2003US-0467199P.  
PR 02-MAY-2003; 2003US-0467201P.  
PR 02-MAY-2003; 2003US-0467203P.  
PR 02-MAY-2003; 2003US-0467230P.  
PR 19-MAY-2003; 2003US-0471306P.  
PR 19-MAY-2003; 2003US-0471336P.  
PR 22-MAY-2003; 2003US-0472420P.  
PR 22-MAY-2003; 2003US-0472430P.  
PR 09-JUN-2003; 2003US-0476609P.  
PR 09-JUN-2003; 2003US-0476621P.  
PR 09-JUN-2003; 2003US-0476632P.  
PR 09-JUN-2003; 2003US-0476641P.  
PR 08-JUL-2003; 2003US-0485217P.  
PR 08-JUL-2003; 2003US-0485218P.  
PR 08-JUL-2003; 2003US-0485223P.  
PR 08-JUL-2003; 2003US-0485224P.  
PR 08-JUL-2003; 2003US-0485325P.  
PR 08-JUL-2003; 2003US-0485359P.  
PR 14-JUL-2003; 2003US-0486446P.  
PR 14-JUL-2003; 2003US-0486480P.  
PR 15-JUL-2003; 2003US-0486891P.  
PR 15-JUL-2003; 2003US-0486960P.  
PR 08-AUG-2003; 2003US-0493341P.  
PR 08-AUG-2003; 2003US-0493370P.  
PR 08-AUG-2003; 2003US-0493573P.  
PR 08-AUG-2003; 2003US-0493577P.  
XX (FIVE-) FIVE PRIME THERAPEUTICS INC.  
XX Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;  
PI Halenbeck RP, Kochakota S, Lin H, Linnemann T, Pierce K, Wang Y;  
PI Wong JGP, Wu G, Zhang H, Zeng C;  
XX WPI; 2004-365511/34.  
XX P-PSDB; AUN99233.  
XX New nucleic acid molecules, useful in preparing a composition for  
PT treating or preventing e.g. inflammatory, CNS, bacterial or viral  
PT disorders, cancer, psoriasis, diabetes, ischemic heart disease or  
PT ulcerative colitis.  
XX Claim 1; SEQ ID NO 49; 532pp; English.  
PS The invention relates to a nucleic acid molecule comprising a  
CC polynucleotide sequence or its complement that encodes a polypeptide. The

CC nucleic acid is useful in preparing a composition for treating or  
CC preventing inflammatory, CNS, immune, bacterial or viral disorder,  
CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic  
CC heart disease or ulcerative colitis. This sequence corresponds to a  
CC nucleic acid of the invention.  
XX  
SQ Sequence 1338 BP; 248 A; 476 C; 345 G; 269 T; 0 U; 0 Other;  
Query Match 30.7%; Score 28.2; DB 12; Length 1338;  
Best Local Similarity 59.3%; Pred. No. 3.7;  
Matches 48; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 10 GAAGACCTCGCGGAGACCTGAAAGCAGCATTGCTCAGATTGCTCCAGTATTACTT 69  
DB 162 GCACGCCCTTTCGACAGCCTGAAGCCAGAGCGTCTCACCTTCCTGCTCCACAC 221  
QY 70 AGCCAGCCCGGTGCTGGGTTT 90  
DB 222 CACAGCCCGCTGGGGAGTT 242  
RESULT 11  
AAS72985  
ID AAS72985 standard; cDNA; 2523 BP.  
XX AAS72985;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #8789.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; as.  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX P-PSDB; ABG08798.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 1; SEQ ID NO 8789; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations

PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225757P.



PR	08-SEP-2000;	2000US-023124441P
PR	08-SEP-2000;	2000US-02314133P
PR	08-SEP-2000;	2000US-02314144P
PR	08-SEP-2000;	2000US-02320801P
PR	08-SEP-2000;	2000US-02320810P
PR	13-SEP-2000;	2000US-02319687P
PR	14-SEP-2000;	2000US-0233977P
PR	14-SEP-2000;	2000US-0233987P
PR	14-SEP-2000;	2000US-0233999P
PR	14-SEP-2000;	2000US-0234000P
PR	14-SEP-2000;	2000US-0234011P
PR	14-SEP-2000;	2000US-02330633P
PR	14-SEP-2000;	2000US-02330644P
PR	21-SEP-2000;	2000US-02330655P
PR	21-SEP-2000;	2000US-02342233P
PR	21-SEP-2000;	2000US-02342744P
PR	25-SEP-2000;	2000US-02349977P
PR	25-SEP-2000;	2000US-02349988P
PR	26-SEP-2000;	2000US-02354844P
PR	27-SEP-2000;	2000US-02356344P
PR	27-SEP-2000;	2000US-02358367P
PR	27-SEP-2000;	2000US-02362377P
PR	29-SEP-2000;	2000US-02363677P
PR	29-SEP-2000;	2000US-02363688P
PR	29-SEP-2000;	2000US-02363699P
PR	29-SEP-2000;	2000US-02363700P
PR	02-OCT-2000;	2000US-02368022P
PR	02-OCT-2000;	2000US-02370377P
PR	02-OCT-2000;	2000US-02370388P
PR	02-OCT-2000;	2000US-02370399P
PR	02-OCT-2000;	2000US-02370400P
PR	13-OCT-2000;	2000US-02399355P
PR	20-OCT-2000;	2000US-02409367P
PR	20-OCT-2000;	2000US-02409368P
PR	20-OCT-2000;	2000US-02412212P
PR	20-OCT-2000;	2000US-02417855P
PR	20-OCT-2000;	2000US-02417866P
PR	20-OCT-2000;	2000US-02417877P
PR	20-OCT-2000;	2000US-02418088P
PR	20-OCT-2000;	2000US-02418099P
PR	20-OCT-2000;	2000US-02418265P
PR	01-NOV-2000;	2000US-02446167P
PR	08-NOV-2000;	2000US-02464745P
PR	08-NOV-2000;	2000US-02464755P
PR	08-NOV-2000;	2000US-02464766P
PR	08-NOV-2000;	2000US-02464777P
PR	08-NOV-2000;	2000US-02464788P
PR	08-NOV-2000;	2000US-02465232P
PR	08-NOV-2000;	2000US-02465243P
PR	08-NOV-2000;	2000US-02465254P
PR	08-NOV-2000;	2000US-02465265P
PR	08-NOV-2000;	2000US-02465276P
PR	08-NOV-2000;	2000US-02465287P
PR	08-NOV-2000;	2000US-02465298P
PR	08-NOV-2000;	2000US-02465309P
PR	08-NOV-2000;	2000US-02466100P
PR	08-NOV-2000;	2000US-02466111P
PR	08-NOV-2000;	2000US-02466122P
PR	17-NOV-2000;	2000US-02492077P
PR	17-NOV-2000;	2000US-02492088P
PR	17-NOV-2000;	2000US-02492099P
PR	17-NOV-2000;	2000US-02492100P
PR	17-NOV-2000;	2000US-02492111P
PR	17-NOV-2000;	2000US-02492122P
PR	17-NOV-2000;	2000US-02492133P
PR	17-NOV-2000;	2000US-02492144P
PR	17-NOV-2000;	2000US-02492155P
PR	17-NOV-2000;	2000US-02492166P
PR	17-NOV-2000;	2000US-02492177P
PR	17-NOV-2000;	2000US-02492188P
PR	17-NOV-2000;	2000US-02492199P
PR	17-NOV-2000;	2000US-02492200P
PR	17-NOV-2000;	2000US-02492211P
PR	17-NOV-2000;	2000US-02492222P
PR	17-NOV-2000;	2000US-02492233P
PR	17-NOV-2000;	2000US-02492244P
PR	17-NOV-2000;	2000US-02492255P
PR	17-NOV-2000;	2000US-02492266P

PR	17-NOV-2000; 2000US-0249255P.
PR	17-NOV-2000; 2000US-0249297P.
PR	17-NOV-2000; 2000US-0249299P.
PR	17-NOV-2000; 2000US-0249300P.
PR	01-DEC-2000; 2000US-0250160P.
PR	01-DEC-2000; 2000US-0250391P.
PR	05-DEC-2000; 2000US-0251030P.
PR	05-DEC-2000; 2000US-0251988P.
PR	05-DEC-2000; 2000US-0256719P.
PR	06-DEC-2000; 2000US-0251479P.
PR	08-DEC-2000; 2000US-0251856P.
PR	08-DEC-2000; 2000US-0251868P.
PR	08-DEC-2000; 2000US-0251869P.
PR	08-DEC-2000; 2000US-0251989P.
PR	08-DEC-2000; 2000US-0251990P.
PR	11-DEC-2000; 2000US-0254097P.
PR	05-JAN-2001; 2001US-0259678P.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Barash SC, Ruben SM;
XX	
XX	WPI; 2001-483426/52..
DR	
XX	
FT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT	useful for preventing, diagnosing and/or treating cancers and metastasis.
XX	
PS	Disclosure; SEQ ID NO 21705; 3071pp + Sequence Listing; English.
XX	
CC	AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC	activity, and can be used in gene therapy and vaccine production. (I)
CC	proteins and polynucleotides may be used in the prevention, diagnosis and
CC	treatment of diseases associated with inappropriate (I) expression. For
CC	example, they may be used to treat disorders associated with decreased
CC	expression by rectifying mutations or deletions in a patient's genome
CC	that affect the activity of (I) by expressing inactive proteins or to
CC	supplement the patients own production of (I). Additionally, (I)
CC	polynucleotides may be used to produce the secreted (I), by inserting the
CC	nucleic acids into a host cell and culturing the cell to express the
CC	protein. (II) proteins and polynucleotides may be used to prevent,
CC	diagnose and treat immune/haematopoietic-related diseases, especially
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC	to AAK87694 represent human immune/haematopoietic antigen genomic
CC	sequences from the present invention. AAK54942 to AAK54950 and AAM92169
CC	represent sequences used in the exemplification of the present invention
XX	
SQ	Sequence 11872 BP; 3363 A; 2615 C; 2631 G; 3263 T; 0 U; 0 Other;
	Query Match            30.7%; Score 28.2; DB 4; Length 11872;
	Best Local Similarity   57.3%; Pred. No. 8.2;
	Matches   51; Conservative   0; Mismatches   38; Indels   0; Gaps   0
Qy	2 GCGATCAGGAACCCCTGGCGGAGAACCTGAAGCACACGATTGCTCATTCGTTCAG 61
Db	4804 GAGATCAAGACCATCCTGGCTTAACACGGTGAACCCGCTTTACTAAAAATACAAAAA 4866
Qy	62 TATTACTTAGCCAGCCGGTGCTGCCTTT 90
Db	4864 TATAGCTGAGCATGGTGGCAGGTGCCCTAT 4892
RESULT 14	
ACA43893	
ID	ACA443893 standard; DNA; 1404 BP.
XX	
XX	ACA43893;
XX	
DT	19-JUN-2003 (first entry)
XX	
DE	Prokaryotic essential gene #25550.
XX	
KW	Antisense; ds: prokaryotic essential gene; cell proliferation;

```

KW drug design; gene.
XX Pseudomonas putida.
OS WO200277183-A2.
XX PN
XX PD
XX PF 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX PA
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX P-PSDB; ABU40023.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 31763; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than S. aureus, S. typhimurium,
XX CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
XX CC prokaryotic essential genes. Note: The sequence data for this patent did
XX CC not form part of the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1404 BP; 328 A; 441 C; 382 G; 253 T; 0 U; 0 Other;
Query Match 30.0%; Score 27.6; DB 8; Length 1404;
Best Local Similarity 56.7%; Pred. No. 6.5;
Matches 51; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1 CGCGATCAGGAAGACCTCGCGGAGAACCTGAAGCAGCAGATTCCTCATTTGCTTCCA 60
DB 1023 CTCGATCCGATTATTCCTTACGTTGGAGCCGGAAGCCCGACGTCATCGAAGCAGCTTCCC 1082
QY 61 GTATTACTTAGCCAGCCGGTGCTGGCTTT 90

Db 1083 GGACCCATCGGCCAACCCGTACCTGGCGTT 1112
RESULT 15
ADR84515/c
ID ADR84515 standard; DNA; 8668 BP.
XX AC ADR84515;
XX
XX 04-NOV-2004 (first entry)
XX DE Aspergillus fumigatus essential gene genomic sequence #326.
XX
XX Fungicide; Aspergillus fumigatus infection; Farmer's lung disease;
XX KW drug screening; ds.
XX
XX Aspergillus fumigatus.
XX OS
XX XX
XX PN WO2004067709-A2.
XX
XX 12-AUG-2004.
XX
XX 16-JAN-2004; 2004WO-US001099.
XX PF
XX 17-JAN-2003; 2003US-0441281P.
XX PR
XX 13-JUN-2003; 2003US-0478196P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX PA (ELIT-) ELITRA CANADA LTD.
XX
XX Jiang B, Hu W, Lemieux S, Roemer T;
XX PI
XX WPI; 2004-594200/57.
XX DR P-PSDB; ADR86276.
XX
XX New purified or isolated Aspergillus fumigatus nucleic acid molecule
XX PT encoding a gene product, useful for diagnosing and/or treating invasive
XX PT fungal infections, such as Farmer's lung disease.
XX
XX Claim 3; SEQ ID NO 326; 164pp; English.
XX
XX The present invention relates to Aspergillus fumigatus genes that are
XX CC essential and are potential targets for drug screening. The methods and
XX CC compositions of the present invention are useful for diagnosing and/or
XX CC treating invasive Aspergillus fumigatus infection, including the allergic
XX CC forms of the disease, such as Farmer's lung disease. They can also be
XX CC used in various drug discovery purposes, such as expression of the
XX CC recombinant protein, hybridization assay and construction of nucleic acid
XX CC arrays. The present sequence represents an Aspergillus fumigatus
XX CC essential gene full length genomic sequence, used during diagnosis and
XX CC drug development in the invention. These genes share a high degree of
XX CC sequence conservation with known essential genes of candida albicans. The
XX CC sequence data for this patent is not represented in the printed
XX CC specification, but was obtained in electronic format from WIPO.
XX
SQ Sequence 8668 BP; 2233 A; 2205 C; 2186 G; 2044 T; 0 U; 0 Other;
Query Match 29.8%; Score 27.4; DB 13; Length 8668;
Best Local Similarity 69.8%; Pred. No. 15;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 4 GATCAGGAAGACCTCGCGGAGAACCTGAAGCAGCAGATTCCTCATTTGCT 56
DB 993 GAATCAATGTCGTCGGGAGACCCGGAACCCAGACCTTCCTCCGTTGCT 941

Search completed: February 16, 2006, 05:20:17
Job time : 342 secs

```

Query Match 100.0%; Score 92; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 4.3e-23;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGATCAGGAAGACCTCGCGAGAACCTGAAAGCAGCATTGCTCATTGCTTCCA 60  
Db 67494 CGCGATCAGGAAGACCTCGCGAGAACCTGAAAGCAGCATTGCTCATTGCTTCCA 67435

Qy 61 GTATTACTTAGCCAGCGGCTGCTGCTTTT 92  
Db 67434 GTATTACTTAGCCAGCGGCTGCTGCTTTT 67403

RESULT 5  
AE005674\_35/c  
WPCOMMENT

Sequence split into 46 fragments LOCUS AE005674 Accession AE005674

Fragment Name	Begin	End
AE005674_00	1	110000
AE005674_01	100001	210000
AE005674_02	200001	310000
AE005674_03	300001	410000
AE005674_04	400001	510000
AE005674_05	500001	610000
AE005674_06	600001	710000
AE005674_07	700001	810000
AE005674_08	800001	910000
AE005674_09	900001	1010000
AE005674_10	1000001	1110000
AE005674_11	1100001	1210000
AE005674_12	1200001	1310000
AE005674_13	1300001	1410000
AE005674_14	1400001	1510000
AE005674_15	1500001	1610000
AE005674_16	1600001	1710000
AE005674_17	1700001	1810000
AE005674_18	1800001	1910000
AE005674_19	1900001	2010000
AE005674_20	2000001	2110000
AE005674_21	2100001	2210000
AE005674_22	2200001	2310000
AE005674_23	2300001	2410000
AE005674_24	2400001	2510000
AE005674_25	2500001	2610000
AE005674_26	2600001	2710000
AE005674_27	2700001	2810000
AE005674_28	2800001	2910000
AE005674_29	2900001	3010000
AE005674_30	3000001	3110000
AE005674_31	3100001	3210000
AE005674_32	3200001	3310000
AE005674_33	3300001	3410000
AE005674_34	3400001	3510000
AE005674_35	3500001	3610000
AE005674_36	3600001	3710000
AE005674_37	3700001	3810000
AE005674_38	3800001	3910000
AE005674_39	3900001	4010000
AE005674_40	4000001	4110000
AE005674_41	4100001	4210000
AE005674_42	4200001	4310000
AE005674_43	4300001	4410000
AE005674_44	4400001	4510000
AE005674_45	4500001	4607203

Continuation (36 of 46) of AE005674 from base 3500001 (AE005674 Shigella flexneri 2a str)

Query Match 100.0%; Score 92; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 4.3e-23;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGATCAGGAAGACCTCGCGAGAACCTGAAAGCAGCATTGCTCATTGCTTCCA 60  
Db 53665 CGCGATCAGGAAGACCTCGCGAGAACCTGAAAGCAGCATTGCTCATTGCTTCCA 53606

Qy 61 GTATTACTTAGCCAGCGGCTGCTGCTTTT 92  
Db 53605 GTATTACTTAGCCAGCGGCTGCTGCTTTT 53574

RESULT 6  
U00096\_35/c  
WPCOMMENT

Sequence split into 47 fragments LOCUS U00096 Accession U00096

Fragment Name	Begin	End
U00096_00	1	110000
U00096_01	100001	210000
U00096_02	200001	310000
U00096_03	300001	410000
U00096_04	400001	510000
U00096_05	500001	610000
U00096_06	600001	710000
U00096_07	700001	810000
U00096_08	800001	910000
U00096_09	900001	1010000
U00096_10	1000001	1110000
U00096_11	1100001	1210000
U00096_12	1200001	1310000
U00096_13	1300001	1410000
U00096_14	1400001	1510000
U00096_15	1500001	1610000
U00096_16	1600001	1710000
U00096_17	1700001	1810000
U00096_18	1800001	1910000
U00096_19	1900001	2010000
U00096_20	2000001	2110000
U00096_21	2100001	2210000
U00096_22	2200001	2310000
U00096_23	2300001	2410000
U00096_24	2400001	2510000
U00096_25	2500001	2610000
U00096_26	2600001	2710000
U00096_27	2700001	2810000
U00096_28	2800001	2910000
U00096_29	2900001	3010000
U00096_30	3000001	3110000
U00096_31	3100001	3210000
U00096_32	3200001	3310000
U00096_33	3300001	3410000
U00096_34	3400001	3510000
U00096_35	3500001	3610000
U00096_36	3600001	3710000
U00096_37	3700001	3810000
U00096_38	3800001	3910000
U00096_39	3900001	4010000
U00096_40	4000001	4110000
U00096_41	4100001	4210000
U00096_42	4200001	4310000
U00096_43	4300001	4410000
U00096_44	4400001	4510000
U00096_45	4500001	4610000
U00096_46	4600001	4639675

Continuation (36 of 47) of U00096 from base 3500001 (U00096 Escherichia coli K-12 MG1655)

Query Match 100.0%; Score 92; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 4.3e-23;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGATCAGGAAGACCTCGCGAGAACCTGAAAGCAGCATTGCTCATTGCTTCCA 60  
Db 75040 CGCGATCAGGAAGACCTCGCGAGAACCTGAAAGCAGCATTGCTCATTGCTTCCA 78981

Qy 61 GTATTACTTAGCCAGCGGCTGCTGCTTTT 92  
Db 78980 GTATTACTTAGCCAGCGGCTGCTGCTTTT 78949

RESULT 7  
BA000007\_42/c  
WPCOMMENT

Sequence split into 55 fragments LOCUS BA000007 Accession BA000007

Fragment Name	Begin	End
BA000007_00	1	110000
BA000007_01	100001	210000
BA000007_02	200001	310000
BA000007_03	300001	410000
BA000007_04	400001	510000
BA000007_05	500001	610000
BA000007_06	600001	710000
BA000007_07	700001	810000
BA000007_08	800001	910000
BA000007_09	900001	1010000
BA000007_10	1000001	1110000
BA000007_11	1100001	1210000
BA000007_12	1200001	1310000
BA000007_13	1300001	1410000
BA000007_14	1400001	1510000
BA000007_15	1500001	1610000
BA000007_16	1600001	1710000
BA000007_17	1700001	1810000
BA000007_18	1800001	1910000
BA000007_19	1900001	2010000
BA000007_20	2000001	2110000
BA000007_21	2100001	2210000
BA000007_22	2200001	2310000
BA000007_23	2300001	2410000
BA000007_24	2400001	2510000
BA000007_25	2500001	2610000
BA000007_26	2600001	2710000
BA000007_27	2700001	2810000
BA000007_28	2800001	2910000
BA000007_29	2900001	3010000
BA000007_30	3000001	3110000
BA000007_31	3100001	3210000
BA000007_32	3200001	3310000
BA000007_33	3300001	3410000
BA000007_34	3400001	3510000
BA000007_35	3500001	3610000
BA000007_36	3600001	3710000
BA000007_37	3700001	3810000
BA000007_38	3800001	3910000
BA000007_39	3900001	4010000
BA000007_40	4000001	4110000
BA000007_41	4100001	4210000
BA000007_42	4200001	4310000
BA000007_43	4300001	4410000
BA000007_44	4400001	4510000
BA000007_45	4500001	4610000
BA000007_46	4600001	4710000
BA000007_47	4700001	4810000
BA000007_48	4800001	4910000
BA000007_49	4900001	5010000
BA000007_50	5000001	5110000
BA000007_51	5100001	5210000
BA000007_52	5200001	5310000
BA000007_53	5300001	5410000
BA000007_54	5400001	5498450

Continuation (43 of 55) of BA000007 from base 4200001 (BA000007 Escherichia coli O157:H7

Query Match 100.0%; Score 92; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 4.3e-23;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CGCGATCAGAGACCCCTCGCGAGAACCTGAAGCAGACATTGCTCATTGCTTCCA 60  
Db 100272 CGCGATCAGAGACCCCTCGCGAGAACCTGAAGCAGACATTGCTCATTGCTTCCA 100213  
Qy 61 GTATTACTTAGCCAGCGGGTGTGCTTTT 92  
Db 100212 GTATTACTTAGCCAGCGGGTGTGCTTTT 100181

RESULT 8  
BA000007\_43/c  
WPCOMMENT

Sequence split into 55 fragments LOCUS BA000007 Accession BA000007

Fragment Name	Begin	End
BA000007_00	1	110000
BA000007_01	100001	210000
BA000007_02	200001	310000
BA000007_03	300001	410000
BA000007_04	400001	510000
BA000007_05	500001	610000
BA000007_06	600001	710000
BA000007_07	700001	810000
BA000007_08	800001	910000
BA000007_09	900001	1010000
BA000007_10	1000001	1110000
BA000007_11	1100001	1210000
BA000007_12	1200001	1310000
BA000007_13	1300001	1410000
BA000007_14	1400001	1510000
BA000007_15	1500001	1610000
BA000007_16	1600001	1710000
BA000007_17	1700001	1810000
BA000007_18	1800001	1910000
BA000007_19	1900001	2010000
BA000007_20	2000001	2110000
BA000007_21	2100001	2210000
BA000007_22	2200001	2310000
BA000007_23	2300001	2410000
BA000007_24	2400001	2510000
BA000007_25	2500001	2610000
BA000007_26	2600001	2710000
BA000007_27	2700001	2810000
BA000007_28	2800001	2910000
BA000007_29	2900001	3010000
BA000007_30	3000001	3110000
BA000007_31	3100001	3210000
BA000007_32	3200001	3310000
BA000007_33	3300001	3410000
BA000007_34	3400001	3510000
BA000007_35	3500001	3610000
BA000007_36	3600001	3710000
BA000007_37	3700001	3810000
BA000007_38	3800001	3910000
BA000007_39	3900001	4010000
BA000007_40	4000001	4110000
BA000007_41	4100001	4210000
BA000007_42	4200001	4310000
BA000007_43	4300001	4410000
BA000007_44	4400001	4510000
BA000007_45	4500001	4610000
BA000007_46	4600001	4710000
BA000007_47	4700001	4810000
BA000007_48	4800001	4910000
BA000007_49	4900001	5010000
BA000007_50	5000001	5110000
BA000007_51	5100001	5210000
BA000007_52	5200001	5310000
BA000007_53	5300001	5410000
BA000007_54	5400001	5498450

Continuation (44 of 55) of BA000007 from base 4300001 (BA000007 Escherichia coli O157:H7

Query Match 100.0%; Score 92; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 4.3e-23;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CGCGATCAGAGACCCCTCGCGAGAACCTGAAGCAGACATTGCTCATTGCTTCCA 60  
Db 272 CGCGATCAGAGACCCCTCGCGAGAACCTGAAGCAGACATTGCTCATTGCTTCCA 213  
Qy 61 GTATTACTTAGCCAGCGGGTGTGCTTTT 92



```

Db      212 GTATTACTAGCCAGCGGCTGCTGCTTTT 181
|||||
RESULT 9
AE016992
LOCUS   Shigella flexneri 2a str. 2457T DNA linear BCT 22-APR-2003
DEFINITION
ACCESSION AE016992
VERSION AE016992.1 GI:30043426
KEYWORDS
SOURCE    Shigella flexneri 2a str. 2457T
ORGANISM  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Shigella.
REFERENCE
AUTHORS   Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
            Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
            Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
            Schwartz,D.C. and Blattner,F.R.
            Complete Genome Sequence and Comparative Genomics of Shigella
            flexneri Serotype 2a Strain 2457T
            Infect. Immun. 71 (5), 2775-2786 (2003)
TITLE
JOURNAL   2 (bases 1 to 289816)
PUBMED    12704152
REFERENCE
AUTHORS   Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
            Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
            Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
            Schwartz,D.C. and Blattner,F.R.
            Direct Submission
            Submitted (13-JUN-2002) Genetics Laboratory, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source    1..289816
            /organism="Shigella flexneri 2a str. 2457T"
            /mol_type="genomic DNA"
            /strain="2457T"
            /serotype="2a"
            /db_xref="taxon:198215"
            complement(238..450)
            /gene="cpa"
            /locus_tag="S4179"
            complement(238..450)
            /genes="cpa"
            /locus_tag="S4179"
            /function="regulator; Adaptations, atypical conditions"
            /notes="residues 1 to 70 of 70 are 100.00 pct identical to
            residues 1 to 70 of 70 from Escherichia coli K-12 : B3556"
            /codon_start=1
            /transl_table=11
            /product="cold shock protein 7.4, transcriptional
            activator of hns"
            /protein_id="AAP19147.1"
            /db_xref="GI:30043427"
            /translation="MSGKMTGIKWNADKGFITPDGSKDVFVHFSAIQNDGYS
            LDGQKVSFTIESGAKGPAACNTSL"
            complement(896..1021)
            /gene="yiaG"
            /locus_tag="S4180"
            complement(896..1021)
            /genes="yiaG"
            /locus_tag="S4180"
            /notes="residues 1 to 41 of 41 are 100.00 pct identical to
            residues 1 to 41 of 96 from Escherichia coli K-12 : B3555"
            /codon_start=1
            /transl_table=11
            /product="hypothetical protein"
            /protein_id="AAP19148.1"
            /db_xref="GI:30043428"
            /translation="MEYKDPMBLLSSLEIQVFKDETKITLTHRTTCTEIBQL"
            1335..2165
            /gene="yiaF"

CDS
1335..2165
/gene="yiaF"
/locus_tag="S4181"
/notes="residues 1 to 276 of 276 are 94.56 pct identical to
residues 1 to 276 of 276 from Escherichia coli K-12 :
B3554"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP19149.1"
/db_xref="GI:30043429"
/translation="MILPGLRRKGILOACPGLSLQSRQTRVCRVCRALFLGRSKMATG-
KSCSRFAPLAALLMVVLSGCGPCDEKGRDOKAFIDPLQNTVMRSGERLPTLTADQKKQ
KGFVSDYAILGYSOQVNOQMDSGLRPVVDSVNAIRVODYVYVTSQGLREMGLGV
LAQLQNAKLQADAAHSALKQSDDLKPFVDQAFKVVTPADALQPLIPAAQTFQQL
VMVDYIAQQGTQVSFVANGIQFTSQASEYNKLIAPLPAQHQAQNAWTAVTATQ
"
complement(2215..3189)
/genes="yiaE"
/locus_tag="S4182"
complement(2215..3189)
/genes="yiaE"
/locus_tag="S4182"
/function="putative enzyme; Not classified"
/notes="residues 1 to 324 of 324 are 94.13 pct identical to
residues 5 to 328 from Escherichia coli K-12 :
B3553"
/codon_start=1
/transl_table=11
/product="putative dehydrogenase"
/protein_id="AAP19150.1"
/db_xref="GI:30043430"
/translation="MKPSVILYKALPDDLQRLQAHFTVHQVANLSQTVVQNAIFA
EAGLLGSNENVDAALEKMKPLRTSTISVGYDNFVDALTARKILLMHTPTVLTET
ADRTLMLVLSTARRVVEVAERKAGEWTASIGPDWYGTDDHHTKILGIVGMGRIGMAL
AQRVHGFNNPILYNARRHHEAEERFNARYCDLTLQESDFVCLILPLTDETHLIF
GAEQPAKSSAIFINAGRGPVVDENALIAALQKGEIHAAGLDVFEQEPVLSVDSPLLS
MANVAVPHIGSSATHTETRYGMACAVDNLIDALQKVKNCVNPHYAD"
complement(3293..3919)
/genes="yiaD"
/locus_tag="S4183"
complement(3293..3919)
/genes="yiaD"
/locus_tag="S4183"
/function="putative membrane; Not classified"
/notes="residues 1 to 208 of 208 are 88.94 pct identical to
residues 12 to 219 of 219 from Escherichia coli K-12 :
B3552"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein"
/protein_id="AAP19151.1"
/db_xref="GI:30043431"
/translation="MSGALAVSGCTTNPYTGERAGKSAIGAGLSLVGAGIGALSSS
KDRGKALIGAAGALGGVGYVMDQEAKLKRDKMRGTGVSVTRSGDNIILNMPNN
VTFDSSAPLKPAGANTLIGVAMVLYKEPKTAVNVIGYTDSTGCHDLNMLRSQRAQS
VASALITQGVDAASRIKRTQGLGPANPIASNSTAGKQNRKRVETLSPL"
4160..6439
/genes="bisC"
/locus_tag="S4184"
4160..6439
/gene="bisC"
/locus_tag="S4184"
/function="enzyme; Biosynthesis of cofactors, carriers;
Biotin"
/notes="residues 21 to 759 of 759 are 98.78 pct identical
to residues 1 to 739 of 739 from Escherichia coli K-12 :
B3551"
/codon_start=1
/transl_table=11
/product="biotin sulfoxide reductase"
/protein_id="AAP19152.1"

```

```
/db_xref="GI:30043432"
/translation="MLVETDGETVFSSRGALATGMENSLOSARVQDVHSTNTRVRPFMV
RKGFLASPNQQRIRGQDFRVSWDEALDLHQHKKRIREAYGPASIPAGSYGWRSN
GVLTKASTPQQIRYALAGAGTYGHLGDYSTGAAQAIMPYVVGSEVYQQQTSWPLVLEH
SDVVLWSANPTLMGLKIANNASDEQSLFSALRDSQKKLICIDPMRSSTVDFDQKM
ENVAEMCTDVALMGIANTLIVENGHDEAFARCTTGAVAFSALYLGSDGIKNAE
WAAEICGVGAAKIRELALPHQNTIMLMAGWQKQOQGEQKHMLITVIAUMLGQIGT
PGGFGSLSYFANGGNPTRRSAVLSSMQSLPGGCDVADKIPIVARIVEALENPGGAYQ
HNGMRHFDPIRPIWAGGANFTHQDNTNRLIRAWQKPELVISCFWTAARAHADIV
LPATTSFERNDLTMTGDYGNQHLVPMKQVPPRYEARNDVDVFAELSERWEGGYARF
TEGKSQLOLETFYNVARQASQOVELPPFAEFQANOLIEMPENPDSEIRFIRADF
CHDPLAHLKTAQSKLEISQRIADYGPCDCHPMWLEPDEWQGNAREQLOVLSAH
PAHRLHSQNLNYSLSRELYAVANREPVIIPDDAQAQKITEGDMVNRWNSRGQLLAGV
ISEG1KPGVICIHEGAWPDLDTADGICKNAGVNLTKDLPSSRLGNCAGNTLAWL
EYKNGPELTLTAFEPASS"
complement (6408..6848)
/gene="yiaC"
/locus_tag="S4185"
complement (6408..6848)
/gene="yiaC"
/locus_tag="S4185"
/notes="residues 1 to 146 of 146 are 97.94 pct identical to
residues 1 to 146 of 146 from Escherichia coli K-12 :
B3550"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP19153.1"
/db_xref="GI:30043433"
/translation="MIREAQRSELPALLEWLESTWGHFFIKANYWRDCIPLVRDAY
LANAQWVNEEGKLLGFSVINEGRPLAAMFVAPKAVRRGIGKALMQYVQQRYPHML
EYVQKNQPAIDFYRAQGFHVDCAWQDETQLPTWMSVPVQTL"
complement (6845..7408)
/gene="tag"
/locus_tag="S4186"
complement (6845..7408)
/gene="tag"
/locus_tag="S4186"
/function="enzyme; DNA - replication, repair,
restriction/modification"
/notes="residues 1 to 187 of 187 are 97.32 pct identical to
residues 1 to 187 of 187 from Escherichia coli K-12 :
B3549"
/codon_start=1
/transl_table=11
/product="3-methyl-adenine DNA glycosylase I"
/protein_id="AAP19154.1"
/db_xref="GI:30043434"
/translation="MERCWVSGGPLYIAYHDNENGVPTDSKKLFEMICFPGQAGL
SWITLKKRNTYRAPHQDPVKVAAQBEDVERLVQDAGIIRHRGKIQAIIIGNARAY
LQMEQNGEPDFVMSFVNHQPVQVQTATLSIPTSTASDALSALKKRGKPFVGT
ICYSFWQACGLVNDHVVGCCCYLGNKP"
7566..8264
/gene="yhjY"
/locus_tag="S4187"
7566..8264
/gene="yhjY"
/locus_tag="S4187"
/function="putative enzyme; Not classified"
/notes="residues 1 to 232 of 232 are 99.13 pct identical to
residues 1 to 232 of 232 from Escherichia coli K-12 :
B3548"
/codon_start=1
/transl_table=11
/product="hypothetical protein yhfL"
/protein_id="AAN82583.1"
/db_xref="GI:26110398"
/translation="NNKFIKVALVCAVLATLTACTGHENRDKNCSYDILLHPAISIS
KIIGCGPTAQ"
complement (432..1517)
/gene="yhfs"
/locus_tag="c4146"
complement (432..1517)
/gene="yhfs"
/locus_tag="c4146"
/notes="Escherichia coli K-12 ortholog: b3376"
/codon_start=1
/transl_table=11
/product="Hypothetical protein yhfS"
/protein_id="AAN82584.1"
/db_xref="GI:26110399"
/translation="MKTFFLOSLTLIEAQKQFALVDTICRHPGCEFLTCDGLGUTP
GLNQPRITQRVEQVLADFAHAQALVQAGTGAIRAAALALKSGRLLVHDAPVYP
TRVYIEQWGLTLITADFNDSALKVQVDEQPDALVQHTRQQQPDQVVLADVLATL
RAAGVPAITDDNYAVMKVARTICEGANNVSTCFKLFQEGVGAVGVGDVSTRRA
TLYSGGSIQGAQALEVLRGLVLPVMAVQAGVSRLLALLNGAVAEVKSAVIANA
KSKRLIVFHFQPIAARVLEBAQKQALPYPVGAESKYEIPPLFYRLSLGSTRQANPOLE
HCAIRINPRNGEETILRLRESIASI"
complement (1529..2833)
/gene="yhft"
/locus_tag="c4147"
complement (1529..2833)
```

```
RESULT 10
AE016768/c
LOCUS
DEFINITION
ACCESSION
VERSION
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1..301660
/organism="Escherichia coli CFT073"
/mol_type="genomic DNA"
/strain="CFT073"
/db_xref="taxon:199310"
157..324
/gene="yhfl"
/locus_tag="c4145"
157..324
/gene="yhfl"
/locus_tag="c4145"
/notes="Escherichia coli K-12 ortholog: b3369; Escherichia
coli O157:H7 ortholog: z4730"
/codon_start=1
/transl_table=11
/product="Hypothetical protein yhfL"
/protein_id="AAN82583.1"
/db_xref="GI:26110398"
/translation="NNKFIKVALVCAVLATLTACTGHENRDKNCSYDILLHPAISIS
KIIGCGPTAQ"
complement (432..1517)
/gene="yhfs"
/locus_tag="c4146"
complement (432..1517)
/gene="yhfs"
/locus_tag="c4146"
/notes="Escherichia coli K-12 ortholog: b3376"
/codon_start=1
/transl_table=11
/product="Hypothetical protein yhfS"
/protein_id="AAN82584.1"
/db_xref="GI:26110399"
/translation="MKTFFLOSLTLIEAQKQFALVDTICRHPGCEFLTCDGLGUTP
GLNQPRITQRVEQVLADFAHAQALVQAGTGAIRAAALALKSGRLLVHDAPVYP
TRVYIEQWGLTLITADFNDSALKVQVDEQPDALVQHTRQQQPDQVVLADVLATL
RAAGVPAITDDNYAVMKVARTICEGANNVSTCFKLFQEGVGAVGVGDVSTRRA
TLYSGGSIQGAQALEVLRGLVLPVMAVQAGVSRLLALLNGAVAEVKSAVIANA
KSKRLIVFHFQPIAARVLEBAQKQALPYPVGAESKYEIPPLFYRLSLGSTRQANPOLE
HCAIRINPRNGEETILRLRESIASI"
complement (1529..2833)
/gene="yhft"
/locus_tag="c4147"
complement (1529..2833)
```

```
Query Match 100.0%; Score 92; DB 1; Length 289816;
Best Local Similarity 100.0%; Pred. No. 4.9e-23;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCGGATCAGAGACCCCTCGGAGAACCTGAAGCAGACATGCTCACATTGCTTCCA 60
DB 135346 CCGGATCAGAGACCCCTCGGAGAACCTGAAGCAGACATGCTCACATTGCTTCCA 135405
QY 61 GTATTACTTAGCCAGCCGGGTGCTGCTTTT 92
DB 135406 GTATTACTTAGCCAGCCGGGTGCTGCTTTT 135437
```

/gene="yhft"  
/locus tag="c4147"  
/function="putative transport"  
/note="Escherichia coli K-12 ortholog: b3377"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein yhfT"  
/protein\_id="AAN82585.1"  
/db\_xref="GI:26110400"

/translation="MDLYIQIIVVACLTGNTSLAHRSAVFHDGIRPILPOLIEGYM  
NRREASIAFGLSIGFVSGISFTIKTGLLNALLFPDIDILGVLAINSLMFAGLGA  
IWGLLITCLLPVQQLTALPVDVLSGLSELSPVSAFALFELVAIFYQGWKQSLI  
LPTLETGLINALGAPGDMQSDSATGWAEIQHEGGDTFMGHQILGTRPLPLRM  
PRFDVDRVEQALVSGWVERGGDDQLFLWVNOVAIGDLEADIGOVYNTANLSV  
ISFDDAIKMGIRVQGVQVITFCGLLTDSCILDAESKEGRGIGINAPRSGAYD  
NGQVVMGTVGVQVQKLYEAGVPTLVGVKVAIDVSNPYGVSWQNLVDSQILMD  
ITLNEFTPTAFICNIQETDLGHAEDVARYAERLQVVDNRNLARLIEAMQDDCLV  
VNADHGNPTIGHSHHTRVPPVLVYQGMVATQLGVRTTLDVGATVCEFFRAPPPQ  
NGRSFLSLRFAGDTL"

gene

complement (2845..3237)

/gene="yhfU"

complement (2845..3237)

/locus tag="c4148"

complement (2845..3237)

/gene="yhfU"

/note="Escherichia coli K-12 ortholog: b3378"

/codon\_start=1

/transl\_table=11

/product="Hypothetical protein yhfU"

/protein\_id="AAN82586.1"

/db\_xref="GI:26110401"

/translation="MKIPLNFSNKRMLKMGVAGLQREQIKKTIATAPGCFEVIH  
NDMEAMKVGSGOLDYVIGACNTGAGAALSATAVIGYNSCTIAKPGIKAKDEHIK  
MIAEGVAFGLSVYVHEHAIPMLINLK"

complement (3209..4087)

/gene="yhfV"

/locus tag="c4149"

complement (3209..4087)

/gene="yhfV"

/locus tag="c4149"

/function="putative"

/note="Escherichia coli K-12 ortholog: b3379"

/codon\_start=1

/transl\_table=11

/product="Phosphodiesterase homology protein"

/protein\_id="AAN82587.1"

/db\_xref="GI:26110402"

/translation="MSFDPTGYTLAHEHLHIDLSPKNNVDCRLDQYAFICQBMNDLM  
ARGVRVIMTRNMGNAQFMLDVNRETGINVACTGYQDAFFPEHVATRSQELA  
QEMVDIEQIDIGTLKAGIIAIEIGSGEKGITSLBEKVFAAALAHNQTGRPISTHTS  
FSTMGLEQALLQAHGVDSLRVTVGHCCLKNDLNMIDLGAVYQVQDTIGKNSYYP  
DEKRIAMHALRDRGLNRLVMSMDITRSHLKAAGGYGVDFLLTTFIQLRQSGFSQ  
ADVDMLENPSQFFQ"

complement (4084..5310)

/gene="yhfW"

/locus tag="c4150"

complement (4084..5310)

/gene="yhfW"

/locus tag="c4150"

/function="putative"

/note="Escherichia coli K-12 ortholog: b3380"

/codon\_start=1

/transl\_table=11

/product="Hypothetical protein yhfW"

/protein\_id="AAN82588.1"

/db\_xref="GI:26110403"

/translation="MARFVLLIDSFVGAMKDVTLVRPDAGANTCGHILSQLPHLQ  
LPTLETGLINALGAPGDMQSDSATGWAEIQHEGGDTFMGHQILGTRPLPLRM  
PRFDVDRVEQALVSGWVERGGDDQLFLWVNOVAIGDLEADIGOVYNTANLSV  
ISFDDAIKMGIRVQGVQVITFCGLLTDSCILDAESKEGRGIGINAPRSGAYD  
NGQVVMGTVGVQVQKLYEAGVPTLVGVKVAIDVSNPYGVSWQNLVDSQILMD  
ITLNEFTPTAFICNIQETDLGHAEDVARYAERLQVVDNRNLARLIEAMQDDCLV  
VNADHGNPTIGHSHHTRVPPVLVYQGMVATQLGVRTTLDVGATVCEFFRAPPPQ  
NGRSFLSLRFAGDTL"

gene

complement (5311..6474)

/gene="yhfX"  
/locus tag="c4151"  
complement (5311..6474)  
/gene="yhfX"  
/locus tag="c4151"  
/note="Escherichia coli K-12 ortholog: b3381"

/codon\_start=1

/transl\_table=11

/product="Hypothetical protein yhfX"

/protein\_id="AAN82589.1"

/db\_xref="GI:26110404"

/translation="MFEVALKRONPALISAALSLWQOQKIAPDSWVIDVQVLENGKR  
GHLVQIPCHOVSDAVEQGTDTVTFTLDKAREISAATKGRVQSLLKVVSDDFLY  
QOESGFVQSHLEHVAETQNLPEGLHAGLTPECLLDDEAAGKVLPTPNLHTLVQAR  
DQAKSGIAIEQNAPSAISCTSLPLAEGVGTHTPEGHALTGTIPANQODQPERIA  
MMLWSISHHFSDSYCGGYGRRGHAQHALVFTPENQRTITYLNVDSDSDYTL  
PLAGEHPVSSAVVLCFRTQIFITRSDVVLVSGIHHGPEIIVGRYDSLGNPLEA"

gene

complement (6558..6962)

/locus tag="c4152"

complement (6558..6962)

/gene="yhfY"

/locus tag="c4152"

/note="Escherichia coli K-12 ortholog: b3382"

/codon\_start=1

/transl\_table=11

/product="Hypothetical protein yhfY"

/protein\_id="AAN82590.1"

/db\_xref="GI:26110405"

/translation="MGNRPVTKRKASDMETRLNLLCDAGVIDKIDCKGMQVNVNLE  
TICHLPARSEQCTMAMTHASALMRSGEEIEPLDELLAELAAQSSHQAVQLHQV  
LLKEFALEVPCEEGYLLANLYGLWMAAEEV"

gene

complement (6937..7785)

/locus tag="c4153"

complement (6937..7785)

/gene="yhfZ"

/locus tag="c4153"

/note="Escherichia coli K-12 ortholog: b3383"

/codon\_start=1

/transl\_table=11

/product="Hypothetical protein yhfZ"

/protein\_id="AAN82591.1"

/db\_xref="GI:26110406"

/translation="MLGKCGNRLKTDIDELATECRSSVGLTQAAKLTLESSGAIRIER  
RGRNSYLVEMDNKALLTHVDINNVCAMPLPYRILYEGLSGLKQAFDGIFFYAHM  
RGADIRVECLLNGVDMVSVRLAAESYLTQKGLALGLGHTYVGEHQLICRKGEN  
ANVKRVGLNRSADQKIMTDVFGSDSDVERVDLSYHESLQRI VKGDVDAVIMWVAEN  
ELTMLGLEATPLTDDPRFLQATEAVILTRVDDYPMQQLRAVVDKHALLAHQQRVVS  
EQEPSY"

gene

complement (8003..9007)

/gene="trpS"

/locus tag="c4154"

complement (8003..9007)

/gene="trpS"

/locus tag="c4154"

Query Match 100.0%; Score 92; DB 1; Length 301660;

Best Local Similarity 100.0%; Pred. No. Se-23;

Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CGCGATCAGAGACCTCGCGAGAACCTGAAAGCAGCATTCATGCTCATTGCTTCCA	60
Db	89615	CGCGATCAGAGACCTCGCGAGAACCTGAAAGCAGCATTCATGCTCATTGCTTCCA	89556

QY	61	GTATTACTTAGCCAGCCGGCTGCTCTTTT	92
Db	89555 <td>GTATTACTTAGCCAGCCGGCTGCTCTTTT <td>89524</td> </td>	GTATTACTTAGCCAGCCGGCTGCTCTTTT <td>89524</td>	89524

RESULT 11

AE017220\_36/c

WPCOMMENT	Sequence split into 48 fragments				LOCUS	AE017220	Accession	AE017220
	Fragment Name	Begin	End					
	AE017220_00	1	110000					
	AE017220_01	100001	210000					
	AE017220_02	200001	310000					
	AE017220_03	300001	410000					
	AE017220_04	400001	510000					
	AE017220_05	500001	610000					
	AE017220_06	600001	710000					
	AE017220_07	700001	810000					
AE017220_08	800001	910000						
AE017220_09	900001	1010000						
AE017220_10	1000001	1110000						
AE017220_11	1100001	1210000						
AE017220_12	1200001	1310000						
AE017220_13	1300001	1410000						
AE017220_14	1400001	1510000						
AE017220_15	1500001	1610000						
AE017220_16	1600001	1710000						
AE017220_17	1700001	1810000						
AE017220_18	1800001	1910000						
AE017220_19	1900001	2010000						
AE017220_20	2000001	2110000						
AE017220_21	2100001	2210000						
AE017220_22	2200001	2310000						
AE017220_23	2300001	2410000						
AE017220_24	2400001	2510000						
AE017220_25	2500001	2610000						
AE017220_26	2600001	2710000						
AE017220_27	2700001	2810000						
AE017220_28	2800001	2910000						
AE017220_29	2900001	3010000						
AE017220_30	3000001	3110000						
AE017220_31	3100001	3210000						
AE017220_32	3200001	3310000						
AE017220_33	3300001	3410000						
AE017220_34	3400001	3510000						
AE017220_35	3500001	3610000						
AE017220_36	3600001	3710000						
AE017220_37	3700001	3810000						
AE017220_38	3800001	3910000						
AE017220_39	3900001	4010000						
AE017220_40	4000001	4110000						
AE017220_41	4100001	4210000						
AE017220_42	4200001	4310000						
AE017220_43	4300001	4410000						
AE017220_44	4400001	4510000						
AE017220_45	4500001	4610000						
AE017220_46	4600001	4710000						
AE017220_47	4700001	4755700						
Continuation (37 of 48) of AE017220 from base 3600001 (AE017220 Salmonella enterica subsp. Typhimurium LT2)								
Query Match 70.4%; Score 64.8; DB 1; Length 110000;								
Best Local Similarity 81.5%; Pred. No. 8.6e-13;								
Matches 75; Conservative 0; Mismatches 17; Indels 0; Gaps 0;								
QY	1	CCGCATCAGAGACCTCGCGAGAACCTGAAAGCAGCACATTGCTCACATTGCTTCCA	60					
Db	68800	CCGCTTCAGGGGACCCCTACGAGAACCTGAAAGCAGCACATTGCTCACATTGCTTCCA	68741					
QY	61	GTATTACTTACGACCGGCTGCTGCTTTT	92					
Db	68740	GTATTATTGGCCAGCTTTTGTGCTGTTTTT	68709					
RESULT 12								
AE008863/c								
LOCUS	Salmonella typhimurium LT2, section 167 of 220 of the complete genome.							
DEFINITION	Salmonella typhimurium LT2, section 167 of 220 of the complete genome.							
ACCESSION	AE008863 AE006468							
VERSION	AE008863.1 GI:16422092							

KEYWORDS	Salmonella typhimurium LT2
SOURCE	Salmonella typhimurium LT2
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE	1 (bases 1 to 23506)
AUTHORS	McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R. and Wilson, R.K.
TITLE	Complete genome sequence of Salmonella enterica serovar Typhimurium LT2
JOURNAL	Nature 413 (6858), 852-856 (2001)
PUBMED	11677609
REFERENCE	2 (bases 1 to 23506)
AUTHORS	The Salmonella typhimurium Genome Sequencing Project
CONSTRM	Direct Submission
TITLE	Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA
JOURNAL	Supported by NIH grant 5U 01 AI43283
COMMENT	Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and GeneMark; http://opal.biology.gatech.edu/GeneMark/
	EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/, and Pedro Romero and Peter Karp at EcoCyc; http://ecocyc.Pangeasystems.com/ecocyc/
	The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset
	This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.
FEATURES	Location/Qualifiers
source	1. .23506
	/organism="Salmonella typhimurium LT2"
	/mol_type="genomic DNA"
	/strain="LT2; SGSC 1412; ATCC 700720"
	/db_xref="ATCC:700720"
	/db_xref="taxon:99287"
	/note="LT2"
	complement(120..2581)
gene	/gene="glgP"
	/note="Synonym: STM3534"
CDS	complement(120..2567)
	/gene="glgP"
	/EC_number="2.4.1.1"
	/note="similar to E. coli glycogen phosphorylase (AAC76453.1); Blastp hit to AAC76453.1 (815 aa), 93% identity in aa 1 - 815"
	/codon_start=1
	/translation="MNAPFTVASPTLSVEALKHSIAYKLMFTIGKDPVTANKHEMLNA TLFAVDRILVERLWRSNAQSQETROYIUSMEFLIGRTLSNALLSGIYDDVKGAL EAMGLDLBELDEENDPGLNGGLGRLAACFLDSLATLGLPGRGYGYDYGFMFKQNI VDGROKESPDYWLKEYGNPWFVKRHNRYKLVFGGRIQOEGKAKRIETEEILAVAYDQ IIPGYDDTATNTRLNNAQASSEINLGNKFNQGDYFAAEDKNHSENVSRVLPDDSTY SGRELRLQKEFLVSATVQDILHRRYQLHKTYENLADKTAIHLNDTHPVLSPILMRL"



CP000026_05	500001	610000
CP000026_06	600001	710000
CP000026_07	700001	810000
CP000026_08	800001	910000
CP000026_09	900001	1010000
CP000026_10	1000001	1110000
CP000026_11	1100001	1210000
CP000026_12	1200001	1310000
CP000026_13	1300001	1410000
CP000026_14	1400001	1510000
CP000026_15	1500001	1610000
CP000026_16	1600001	1710000
CP000026_17	1700001	1810000
CP000026_18	1800001	1910000
CP000026_19	1900001	2010000
CP000026_20	2000001	2110000
CP000026_21	2100001	2210000
CP000026_22	2200001	2310000
CP000026_23	2300001	2410000
CP000026_24	2400001	2510000
CP000026_25	2500001	2610000
CP000026_26	2600001	2710000
CP000026_27	2700001	2810000
CP000026_28	2800001	2910000
CP000026_29	2900001	3010000
CP000026_30	3000001	3110000
CP000026_31	3100001	3210000
CP000026_32	3200001	3310000
CP000026_33	3300001	3410000
CP000026_34	3400001	3510000
CP000026_35	3500001	3610000
CP000026_36	3600001	3710000
CP000026_37	3700001	3810000
CP000026_38	3800001	3910000
CP000026_39	3900001	4010000
CP000026_40	4000001	4110000
CP000026_41	4100001	4210000
CP000026_42	4200001	4310000
CP000026_43	4300001	4410000
CP000026_44	4400001	4510000
CP000026_45	4500001	4595229
Continuation (35 of 46) of CP000026 from base 3400001 (CP000026 Salmonella enterica subsp		
Query Match 69.3%; Score 63.8; DB 1; Length 110000;		
Best Local Similarity 81.3%; Pred. No. 2.1e-12;		
Matches 74; Conservative 0; Mismatches 17; Indels 0; Gaps 0;		
QY	2	CGCATCAGGAGACCTCGCGGAGAACCTGAAAGCAGCAGCATTTGCTCACAATTGCTTCCAG 61
DB	106615	GCATTAGGGGACCCCTACGGAGAACCTGAAAGCAGCAGCATTTGCTCACAATTGCTTCCAG 106556
QY	62	TATTACTTAGCCAGCGGTGCTGCTTTT 92
DB	106555	TATTATTGGCAGCTTTTGTGCTTTT 106525
RESULT 14		
CP000026_35/c		
WPCOMMENT		
Sequence split into 46 fragments LOCUS CP000026 Accession CP0000026		
Fragment Name	Begin	End
CP000026_00	1	110000
CP000026_01	100001	210000
CP000026_02	200001	310000
CP000026_03	300001	410000
CP000026_04	400001	510000
CP000026_05	500001	610000
CP000026_06	600001	710000
CP000026_07	700001	810000
CP000026_08	800001	910000
CP000026_09	900001	1010000
CP000026_10	1000001	1110000
CP000026_11	1100001	1210000

CP000026_12	1200001	1310000
CP000026_13	1300001	1410000
CP000026_14	1400001	1510000
CP000026_15	1500001	1610000
CP000026_16	1600001	1710000
CP000026_17	1700001	1810000
CP000026_18	1800001	1910000
CP000026_19	1900001	2010000
CP000026_20	2000001	2110000
CP000026_21	2100001	2210000
CP000026_22	2200001	2310000
CP000026_23	2300001	2410000
CP000026_24	2400001	2510000
CP000026_25	2500001	2610000
CP000026_26	2600001	2710000
CP000026_27	2700001	2810000
CP000026_28	2800001	2910000
CP000026_29	2900001	3010000
CP000026_30	3000001	3110000
CP000026_31	3100001	3210000
CP000026_32	3200001	3310000
CP000026_33	3300001	3410000
CP000026_34	3400001	3510000
CP000026_35	3500001	3610000
CP000026_36	3600001	3710000
CP000026_37	3700001	3810000
CP000026_38	3800001	3910000
CP000026_39	3900001	4010000
CP000026_40	4000001	4110000
CP000026_41	4100001	4210000
CP000026_42	4200001	4310000
CP000026_43	4300001	4410000
CP000026_44	4400001	4510000
CP000026_45	4500001	4585229
Continuation (36 of 46) of CP000026 from base 3500001 (CP000026 Salmonella enterica subsp		
Query Match 69.3%; Score 63.8; DB 1; Length 110000;		
Best Local Similarity 81.3%; Pred. No. 2.1e-12;		
Matches 74; Conservative 0; Mismatches 17; Indels 0; Gaps 0;		
QY	2	CGCATCAGGAGACCTCGCGGAGAACCTGAAAGCAGCAGCATTTGCTCACAATTGCTTCCAG 61
DB	6615	GCATTAGGGGACCCCTACGGAGAACCTGAAAGCAGCAGCATTTGCTCACAATTGCTTCCAG 6556
QY	62	TATTACTTAGCCAGCGGTGCTGCTTTT 92
DB	6555	TATTATTGGCAGCTTTTGTGCTTTT 6525
RESULT 15		
AL627281		
LOCUS		
DEFINITION		
Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,		
complete chromosome; segment 17/20.		
ACCESSION		
AL627281 AL513382		
VERSION		
AL627281.1 GI:16504930		
KEYWORDS		
SOURCE		
ORGANISM		
Salmonella enterica subsp. enterica serovar Typhi		
Salmonella enterica subsp. enterica serovar Typhi		
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
Enterobacteriaceae; Salmonella.		
REFERENCE		
1 (bases 1 to 285050)		
AUTHORS		
Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,		
Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.,		
Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,		
Conerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,		
Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,		
Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moulis,S., O'Gaora,P.,		
Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,		
Stevens,K., Whitehead,S. and Barrell,B.G.		
Complete genome sequence of a multiple drug resistant Salmonella		
enterica serovar Typhi CT18		
JOURNAL		
Nature 413 (6858), 848-852 (2001)		



```
/gene="STY4115"
/notes="Pfam match to entry PF00480 ROK, ROK family, score
122.90, E-value 1.8e-35"
complement(5424..5427)
/notes="possible RBS"
5573..5579
/notes="possible RBS"
5586..7124
/genes="STY4116"
/notes="synonym: aldB"
5586..7124
/genes="STY4116"
/EC_number="1.2.1.22"
/notes="Fasta hit to ALDA_ECOLI (478 aa), 33% identity in
478 aa overlap
Fasta hit to FEAB_ECOLI (499 aa), 34% identity in 481 aa
overlap
Fasta hit to YNEI_ECOLI (462 aa), 31% identity in 465 aa
overlap
Fasta hit to DHAB_ECOLI (489 aa), 38% identity in 478 aa
overlap
Fasta hit to DHAL_ECOLI (495 aa), 40% identity in 481 aa
overlap
Fasta hit to YDCW_ECOLI (474 aa), 35% identity in 481 aa
overlap
Fasta hit to GABD_ECOLI (482 aa), 34% identity in 485 aa
overlap
Orthologue of E. coli aldB (ALDB_ECOLI); Fasta hit to
ALDB_ECOLI (512 aa), 96% identity in 512 aa overlap"
/codon_start=1
/translation=11
/product="aldehyde dehydrogenase B"

Query Match      64.1%; Score 59; DB 1; Length 265050;
Best Local Similarity 78.0%; Pred. No. 1.5e-10;
Matches 71; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY      2  GCGATCAGGAAGACCTCGCGGAGAACCTGAAGACGACGACATTGCTCACATTGCTTCCAG 61
Db      157180  GCATTCAGGGGAACCCCTACGGATAACCTGAAGACGACGACATTGCTCACATTGCTTCCAG 157239

QY      62  TATTACTTAGCCAGCGGGTCTGGCTTTT 92
Db      157240  TTTATTTGGCAGCTTTTGTGCTTTT 157270

Search completed: February 16, 2006, 05:50:26
Job time : 1806 secs
```